PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/29, 15/57, 15/82, A01H 5/00	A2	(11) International Publication Number: **WO 97/35983**
		(43) International Publication Date: 2 October 1997 (02.10.97)

(21) International Application Number: PCT/GB97/00729

(22) International Filing Date: 18 March 1997 (18.03.97)

(30) Priority Data:
9606062.9 22 March 1996 (22.03.96) GB

(71) Applicant *(for all designated States except US)*: ZENECA LIMITED [GB/GB]; 15 Stanhope Gate, London W1Y 6LN (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: JEPSON, Ian [GB/GB]; Jealott's Hill Research Station, Bracknell, Berkshire RG42 6EY (GB). GREENLAND, Andrew, James [GB/GB]; Jealott's Hill Research Station, Bracknell, Berkshire RG42 6EY (GB). THOMAS, Didier, Rene, Philippe [BE/GB]; Jealott's Hill Research Station, Bracknell, Berkshire RG42 6EY (GB).

(74) Agents: KENT, Lindsey, Ruth et al.; Zeneca Agrochemicals, Intellectual Property Dept., Jealott's Hill Research Station, P.O. Box 3538, Bracknell, Berkshire RG42 6YA (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: CYSTEINE PROTEASE PROMOTER FROM OIL SEED RAPE AND A METHOD FOR THE CONTAINMENT OF PLANT GERMPLASM

(57) Abstract

A promoter comprising the DNA sequence of an oil seed rape cysteine protease gene promoter of class 1, 2 or 6 is described. The promoter may be used in an expression system for at least the tissue or tissues of a germinating seedling or developing grain or plant (e.g. in the root, cotyledons, leaves and stem). In a preferred embodiment, the expression system comprises a disrupter gene fused to a promoter according to the present invention.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

CYSTEINE PROTEASE PROMOTER FROM OIL SEED RAPE AND A METHOD FOR THE CONTAINMENT OF PLANT GERMPLASM

The present invention relates to promoters and to a construct comprising the same. The present invention also relates to a method for the containment of plant germplasm.

In particular, the present invention relates to the use of a promoter for the expression of a gene of interest (GOI) in a specific tissue or tissues of a plant.

More particularly, the present invention relates to promoters for cysteine proteases. The present invention also relates to the application of these cysteine protease promoters to express a GOI in a specific tissue or tissues of a plant.

Promoters control the spatial and temporal expression of genes by modulating their level of transcription. Early approaches to genetically engineered crop plants utilised strong constitutive promoters to drive the expression of foreign genes. As strategies in plant biotechnology have become more sophisticated, there are requirements for specific promoters to target transgene expression to a particular tissue or to a particular developmental stage.

Cysteine proteases are members of a large multigene family in plants (Praekelt *et al.*, 1988; Goetting-Minesky and Mullin, 1994), animals (Wiederanders *et al.*, 1992) and protozoa (Mallinson *et al*, 1994). Cysteine proteases are synthesised as an inactive precursor (Praekelt *et al.*, 1988). The pre-pro-enzyme is targeted to the secretory pathway (Marttila *et al.*,1995) and post-transcriptionally processed in the vacuoles by proteolytic cleavage of the propeptide fragment to produce the active enzyme (Hara-Nishimura *et al.*, 1993 and 1994).

Plant cysteine proteases participate in different metabolic events of physiological importance. During seed germination and plant senescence they are involved in protein degradation (Jones *et al.*, 1995; Valpuesta *et al.*, 1995; Smart *et al.*, 1995) and play a key role in protein storage mobilisation during germination (Boylan and Sussex, 1987). During seed development, cysteine proteases catalyse the post-translational processing of protein precursors into their mature form (Hara-Nishimura *et al*, 1995). In addition, some are subjected to hormonal regulation either by giberellic acid (Koehler and Ho, 1990; Watanabe *et al.*, 1991) or ethylene (Cervantes *et al.*, 1994; Jones *et al.*, 1995). Others are induced in response to stress like wounding (Linthorst *et al.*, 1993; Lidgett *et al.*, 1995), dehydration (Guerrero *et al.*, 1990), cold (Schaffer and Fischer, 1988) or are implicated in plant-microbe interactions (Goetting-Minesky and Mullin, 1994).

Germination specific cysteine proteases have been characterised for barley (Marttila *et al.*, 1995), rice (Watanabe *et al.*, 1991), maize (Debarros and Larkins, 1994), chick-pea (Cervantes *et al.*, 1994), vetch (Becker *et al*, 1994) and a cysteine protease has been described for oil seed rape (Comai and Harada, 1989). However, the published data for oil seed rape is contradictory. Furthermore, this species is difficult to study due to its amphi-diploid nature. Rather than using more conventional and laborious techniques like subtractive or differential screening of cDNA libraries or differential display techniques, potentially generating clones of unknown identity, cysteine proteinases (cysteine proteases) in oil seed rape were studied which are expressed in germinating seeds. Promoters from genes which are uniquely expressed following seed germination were isolated and characterised.

Thus, according to a first aspect of the present invention, there is provided an oil seed rape cysteine protease gene promoter of class 1, 2 or 6.

According to a second aspect of the present invention, there is provided a promoter comprising at least part of a sequence as shown in Figures 19, 20 or 21, or at least part of a sequence that has substantial homology therewith, or a variant thereof.

According to a third aspect of the present invention, there is provided a promoter having the characteristic motifs or features of promoters of the present invention.

According to a fourth aspect of the present invention, there is provided a recombinant DNA construct comprising the promoter as defined above operably linked to a gene which codes for a protein of interest.

According to a fifth aspect of the present invention, there is provided a recombinant DNA construct functional in a plant comprising a disrupter gene encoding a product capable of disrupting cell function, and a promoter as defined above, the disrupter gene being functionally linked to and controlled by an externally regulatable gene control region which includes a promoter which is inducible by the external application of a chemical inducer.

According to a sixth aspect of the present invention, there is provided DNA comprising at least part of the sequence shown in Figures 12, 13, 14, 15, 16 or 17, or at least part of a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.

According to a seventh aspect of the present invention, there is provided a recombinant DNA construct functional in a plant comprising the DNA as defined above operably linked to a promoter.

According to an eighth aspect of the present invention, there is provided an expression system for the tissue or tissues of a plant material, the expression system comprising a gene of interest fused to a gene promoter as defined above wherein the expression system is capable of being expressed in the tissue or tissues of the plant material.

According to a ninth aspect of the present invention, there is provided an expression system comprising a construct as defined above.

According to a tenth aspect of the present invention, there is provided a recombinant plant genome comprising a promoter as defined above, DNA as defined above, a recombinant DNA construct as defined above or an expression system as defined above.

According to an eleventh aspect of the present invention, there is provided a plant, plant seed or plant cell having a recombinant plant genome as defined above.

According to a twelfth aspect of the present invention, there is provided protected germplasm comprising a recombinant DNA construct as defined above.

According to a thirteenth aspect of the present invention, there is provided a plant or seed which is capable of growing to maturity comprising a recombinant DNA construct as defined above.

According to a fourteenth aspect of the present invention, there is provided the use of a gene promoter as defined above to induce expression of a gene of interest when fused to the gene promoter in the tissue or tissues of a plant material.

Preferably, the inducible promoter of the recombinant DNA construct is functionally linked to and controls a repressor protein and the disrupter gene promoter includes an operator sequence which is recognised by the repressor protein, so that in the presence of the inducer the repressor protein is produced which interacts with the operator sequence thereby disabling the second promoter and inhibiting expression of the disrupter gene.

Preferably, the disrupter gene is a nucleotide sequence, which is in sense orientation to an endogenous plant gene which is essential to plant development or a gene conferring a

desired characteristic on the plant, or comprises a partial sense sequence of the endogenous plant gene.

Preferably, the disrupter gene is a nucleotide sequence which is in antisense orientation to an endogenous plant gene which is essential to plant development or a gene conferring a desired characteristic on the plant.

Preferably, the endogenous plant gene is essential to seed germination or early seedling development.

Preferably, the externally regulatable gene control region is a chemically inducible gene promoter sequence from the glutathione S-transferase system (which is the subject of our International Patent Application No. PCT/ GB96/02116), the *Alc* system (which is the subject of our International Patent Application Nos. PCT/GB96/01883 and PCT/GB96/01846) or the ecdysone system (which is the subject of our International Patent Application No. PCT/GB96/01195).

Preferably, the repressor protein gene encodes a bacterial repressor such as the *lac* repressor or a repressor used by 434, P22 or lambda-bacteriophages.

Preferably, the disrupter gene or disrupter promoter contains a "pseudo-operator".

Preferably, the disrupter gene is a cytotoxic gene.

Preferably, the disrupter gene encodes a recombinase or a transposase adapted to excise a nucleotide sequence flanked by recombinase recognition sequences.

Preferably, the recombinant DNA construct is capable of being expressed in the tissue or tissues of a germinating seedling or a developing grain or a plant when the construct is integrated, preferably stably integrated, within the grain's or seedling's or plant's genomic DNA.

Preferably, the expression system is for at least the tissue of a germinating seedling or developing grain or plant (eg in the root, cotyledons, leaves and stem).

Preferably, the expression system is integrated, preferably stably integrated, within a germinating seedling's genomic DNA or a developing grain's genomic DNA or a plant's genomic DNA.

Preferably, the gene promoter is used to induce expression of a gene of interest when fused to the gene promoter in at least the tissue or tissues of a germinating seedling or a developing grain or a plant (eg in the root, cotyledons, leaves and stem).

According to a preferred embodiment of the present invention, the promoter comprises a DNA sequence corresponding to that of the promoter region of the clone pKS12p6, as shown in Figure 19.

According to another preferred embodiment of the present invention, the promoter comprises a DNA sequence corresponding to that of the promoter region of the clone pKS25p7, as shown in Figure 20.

According to a further preferred embodiment of the present invention, the promoter comprises a DNA sequence corresponding to that of the promoter region of the clone pKS66p1, as shown in Figure 21.

An even more preferred embodiment of the present invention is a seedling, grain or plant comprising a construct comprising a disrupter gene fused to a cysteine protease promoter, wherein the construct is integrated, preferably stably integrated within the seedling's, grain's or plant's genomic DNA, wherein the promoter comprises at least part of a sequence shown in Figures 19, 20 or 21, or at least part of a sequence that has substantial homology therewith, or a variant thereof and wherein the disrupter gene is a gene which encodes barnase ribonuclease.

An even more preferred embodiment of the present invention is a seedling, grain or plant comprising a construct comprising a disrupter gene fused to a cysteine protease promoter, wherein the construct is integrated, preferably stably integrated within the seedling's, grain's or plant's genomic DNA, wherein the promoter comprises at least part of a sequence shown in Figures 19, 20 or 21, or at least part of a sequence that has substantial homology therewith, or a variant thereof and wherein the disrupter gene is a gene which encodes a recombinase adapted to excise a nucleotide sequence flanked by recombinase recognition sequences.

Thus, according to a highly preferred embodiment of the present invention, there is provided a recombinant DNA construct for insertion into the genome of a plant to impart control of plant development thereto, comprises, in sequence:

(a) an inducible gene promoter sequence responsive to the presence or absence of an exogenous chemical inducer;

(b) either a gene encoding a repressor protein under control of the said inducible gene promoter sequence or a gene encoding an inhibitor of the product of the disrupter gene specified at (e) below;

(c) an operator sequence responsive to the said repressor protein;

(d) a gene promoter sequence of the present invention; and,

(e) a gene encoding a protein disrupter of a plant characteristic essential to the growth of the plant, whereby the presence or absence of the exogenous chemical inducer enables either growth to maturity or causes growth to slow down or stop at an appropriate stage.

The inducible promoter used in the present invention may promote expression of the repressor protein in response to stimulation by an exogenous chemical inducer whereby in the absence of the chemical inducer no repressor protein is expressed to interact with the operator thus permitting expression of the disrupter protein gene and in the presence of the chemical inducer repressor protein is expressed thereby preventing expression of the gene encoding the inhibitor of plant development permitting unimpeded plant growth.

The term "plant material" includes a developing caryopsis , a germinating caryopsis or grain, or a seedling, a plantlet or a plant, or tissues or cells thereof, such as the cells of a developing caryopsis or the tissues of a germinating seedling or developing grain or plant (eg in the root, leaves and stem).

The term "gene of interest" or "GOI" with reference to the present invention means any gene of interest. A GOI can be any gene that is either foreign or natural to the plant in question, except for the wild type functional gene when in its natural environment.

Typical examples of a GOI includes genes encoding for proteins and enzymes that disrupt cell function. For example, the gene may be a cytotoxic gene. Alternatively the gene may encode a recombinase, transposase, or a related enzyme with similar properties, adapted to inhibit an endogenous plant gene which is essential to plant development or a gene conferring a desired characteristic on the plant.

A recombinase is an enzyme that recognises a specific excision sequence or set of specific excision sequences and effects the removal of, or otherwise alters, DNA between

specific excision sequences. Recombinase systems such as the Cre-lox, the FLP, SR1 and SSV1-encoded integrase systems may be used in the present invention.

Other examples of a GOI include defensive or protective genes, such as genes giving herbicide, fungal or insect resistance. Such genes may be expressed during germination of seedlings at which time the seedlings are particularly vulnerable. Preferably, the gene encodes a protein which confers resistance to biotic and environmental stresses on a plant.

Also included are endogenous genes such as genes encoding β-tubulin and adenine nucleotide translocator (ANT).

The term "disrupter gene" is a gene which, when expressed or repressed specifically at a suitable stage of plant development, will lead to the failure of a plant to reach maturity and to set seed. The origin of the disrupter genes can be from a variety of naturally occurring sources eg human cells, bacterial cells, yeast cells, plant cells, fungal cells, or they can be totally synthetic genes which may be composed of DNA sequences, some of which may be found in nature, some of which are not normally found in nature or a mixture of both. The disrupter genes will preferably be targeted to an essential biochemical function, such as DNA and RNA metabolism, protein synthesis, and other metabolic pathways.

In a preferred embodiment, the disrupter gene is a gene which encodes barnase ribonuclease, β-tubulin or adenine nucleotide translocator (ANT).

The term "variant thereof" with reference to the present invention means any substitution of, variation of, modification of, replacement of, deletion of or the addition of one or more nucleic acid(s) from or to the promoter sequence providing the resultant sequence is capable of expressing a GOI. The term also includes sequences that can substantially hybridise to the promoter sequence. The term also includes DNA which hybridises to the DNA of the present invention and which codes for at least part of a cysteine protease promoter. Preferably, such hybridisation occurs at, or between, low and high stringency conditions. In general terms, low stringency conditions can be defined as 3 x SSC at about ambient temperature to about 65°C, and high stringency conditions as 0.1 x SSC at about 65°C. SSC is the name of a buffer of 0.15M NaCl, 0.015M trisodium citrate. 3 x SSC is three time as strong as SSC and so on.

The term "substantial homology" covers homology with respect to at least the essential nucleic acids/nucleic acid residues of the promoter sequence providing the

homologous sequence acts as a promoter eg a cysteine protease promoter which is capable of expressing a GOI. Typically, homology is shown when 60% or more of the nucleotides are common with the promoter sequence of the present invention, more typically 65%, preferably 70%, more preferably 75%, even more preferably 80% or 85% and, especially preferred, are 90%, 95%, 98% or 99% or more homology.

The term "construct" - which is synonymous with terms such as "cassette". "hybrid" and "conjugate" - includes a GOI directly or indirectly attached to the promoter of the present invention, such as to form a cassette. An example of an indirect attachment is the provision of a suitable spacer group such as an intron sequence, intermediate the promoter and the GOI. The same is true for the term "fused" in relation to the present invention which includes direct or indirect attachment.

The term "expression system" means that the system defined above can be expressed in an appropriate organism, tissue, cell or medium. In this regard, the expression system of the present invention may comprise additional components that ensure to increase the expression of the GOI by use of the gene promoter.

The DNA of the present invention may be genomic DNA which is in an isolated form and is, preferably, operably linked to a sequence with which it is not naturally associated, or the DNA may be synthetic DNA or cDNA.

Young seedlings at germination represent a vulnerable stage of plant development. Strategies to improve crop production may include expression of genes during this stage to enhance the seedling's resistance to biotic and environmental stresses such as resistance to cold, salt, heavy metals, fungal attack. Protecting seedlings by expressing proteins which provide at least some measure of resistance/tolerance against such stresses, for example anti-fungal proteins (Cammue *et al.*, 1992; Terras *et al.*, 1993) under the control of a germination specific promoter will limit the expression to a precise phase of development, when the proteins will be most effective. Such development specific expression has further advantages such as avoiding expression of these genes in plant material entering the food chain.

The promoters of the present invention may also be advantageously used in plant germplasm containment systems.

Agriculture uses many crop plants for the production of food for human consumption, for commercial processes yielding products for human consumption, for animal feedstuff

production, for the development of industrial products and other purposes. The process involves the planting by the farmer of seed which usually has been purchased from a seed producer. The product produced by the crop, be it the whole plant, the seed or fruit of the plant, is harvested and is then used for the various food applications mentioned above.

The supplied hybrid or inbred seed may incorporate novel genetic information introduced by transformation of the crop giving novel agronomic features such as tolerance to herbicides, insect pests, and fungal diseases, improved yield and/or quality of the harvested product, and novel mechanisms for the control of plant fertility. Such improvements which are made possible through biotechnological research, improve the quality of the plant breeding and improve the agronomic performance of the seed supplied to the farmer.

A problem addressed by the present invention is the containment of crop plants within the area of cultivation. Seeds of cultivated crop plants may be conveyed outside the defined growing area by a number of routes (by birds or small mammals or simply by being dropped during post-harvest transport of a seed crop) where they assume the status of weeds, or they may remain as volunteers in a subsequent crop in later years. It would clearly be appropriate, if it were possible, that cultivated crops be confined to the growing area and prevented from persisting in the wild. It will be appreciated that the problems of crop non-confinement mentioned above become more acute where transgenic crops are involved.

In the same way, pollen can travel long distances through wind and/or insect transportation (dispersion) and remain viable for a long period. Since interspecific crossing between crop plants and their non cultivated related species is possible, a second concern is the escape of pollen from transgenic crops e.g. herbicide resistant crops, to their related weeds species (Mikkelsen et al., 1996). Ways to reduce viability of such hybrids would limit the risk of transgene escape to non-crop species thus avoiding the spreading of plants with enhanced invasiveness or weediness.

It will be appreciated that the use of the seedling promoters of the present invention restricts expression of the disrupter protein gene to a suitable stage of plant development, and also means that it is not necessary to continue to apply an inducer chemical to the plant throughout its lifetime in order to maintain its viability. This has both economic and ecological benefits.

The invention also provides a genetically transformed plant and parts thereof, such as cells protoplasts and seeds, having incorporated, preferably stably incorporated, into the genome the construct of the present invention.

Thus, the invention provides a plant which can be reversibly inhibited at an appropriate developmental stage in which said plant contains, preferably stably incorporated in its genome, the recombinant DNA construct defined above.

Expression of a protein encoded by a gene is controlled by the interaction of certain regulatory proteins, known as DNA-binding proteins, with a region located upstream of the gene. Within the promoter region, there are located several operator regions which contain a specific oligonucleotide sequence to which these DNA-binding proteins specifically bind. These proteins can lead either to activation or repression of gene expression. Thus, they control the regulated expression of genes.

These DNA-binding proteins, which may in fact be either repressors or activators of gene expression, are herein referred to for the sake of simplicity as "repressors".

The present invention makes use of the well-characterised interaction between bacterial operators with their repressors to control the expression of the disrupter gene function. Bacterial repressors, particularly the *lac* repressor, or repressors used by 434, P22 and lambda bacteriophages can be used to control the expression in plant cells very effectively.

A second operator/repressor system is the subject of our published International Patent Application No. WO90/08827 which is incorporated herein by reference.

A third approach for the down-regulation of the disrupter genes which can be considered is the use of either "antisense", "sense" or "partial sense" technology.

A DNA construct according to the invention may be an "antisense" construct generating "antisense" RNA or a "partial-sense" construct (encoding at least part of the functional gene product) generating "sense" RNA.

"Antisense RNA" is an RNA sequence which is complementary to a sequence of bases in the corresponding mRNA: complementary in the sense that each base (or the majority of bases) in the antisense sequence (read in the 3' to 5' sense) is capable of pairing with the corresponding base (G with C, A with U) in the mRNA sequence read in the 5' to 3' sense.

Such antisense RNA may be produced in the cell by transformation with an appropriate DNA construct arranged to generate a transcript with at least part of its sequence complementary to at least part of the coding strand of the relevant gene (or of a DNA sequence showing substantial homology therewith).

"Sense RNA" is an RNA sequence which is substantially homologous to at least part of the corresponding mRNA sequence. Such sense RNA, or partial sense RNA, may be produced in the cell by transformation with an appropriate DNA construct arranged in the normal orientation so as to generate a transcript with a sequence identical to at least part of the coding strand of the relevant gene (or of a DNA sequence showing substantial homology therewith). Suitable sense, or partial sense constructs may be used to inhibit gene expression (as described in International Patent Publication WO91/08299).

Antisense RNA constructs which may be used to down-regulate disrupter genes include those encoding adenine nucleotide translocator.

Other approaches which are, or become, available may also be used.

Further details on such crop containment systems can be found in our published International Patent Application No. WO94/03619 which is incorporated herein by reference.

The promoter of the present invention, then, when linked to an exogenous or foreign gene and introduced into a plant by transformation, provides a means for the regulation of expression of that foreign gene. The method employed for transformation of the plant cells is not especially germane to this invention and any method suitable for the target plant may be employed. Transgenic plants are obtained by regeneration from the transformed cells. Numerous transformation procedures are known from the literature such as agroinfection using *Agrobacterium tumefaciens* or its Ti plasmid, electroporation, microinjection or plants cells and protoplasts, microprojectile transformation, to mention but a few. Reference may be made to the literature for full details of the known methods.

Neither is the plant species into which the promoter sequence is inserted particularly germane to the invention. Dicotyledonous and monocotyledonous plants can be transformed. This invention may be applied to any plant for which transformation techniques are, or become, available. The present invention can therefore be used to control gene expression in a variety of genetically modified plants, including field crops such as canola, sunflower,

tobacco, sugarbeet, and cotton; cereals such as wheat, barley, rice, maize, and sorghum; fruit such as tomatoes, mangoes, peaches, apples, pears, strawberries, bananas and melons; and vegetables such as carrot, lettuce, cabbage and onion. The promoter is also suitable for use in a variety of tissues, including roots, leaves, stems and reproductive tissues.

Thus, nucleic acid sequences which code for novel cysteine proteases have been isolated and characterised. The DNA comprising at least part of the sequence shown in any one of Figures 12 to 17 codes for a cysteine protease and corresponds to the coding region of the sequence. The present invention also includes DNA which shows homology to the sequences of the present invention. The present invention also includes DNA which hybridises to the DNA of the present invention and which codes for at least part of a cysteine protease. Such homology and hybridisation is discussed above in relation to the promoter sequences.

The present invention further includes DNA which is degenerate as a result of the genetic code to the DNA of the present invention and which codes for a cysteine protease.

Also provided by the present invention is a cysteine protease, which is substantially free from other proteins with which it is ordinarily associated, and which is coded for by cysteine protease gene DNA of the present invention.

The present invention will now be described by way of non-limiting example only, and with references to the accompanying drawings, in which:

Figure 1 shows a schematic outline of the identification of cysteine protease isoforms using a reverse transcribed PCR library;

Figure 2 shows the analysis of RT-PCR products by agrose gel electrophoresis and in which DG1 = DEGCYS1 and DG2 = DEGCYS2 oligos;

Figure 3A shows the alignment of the coding regions of the preliminary nucleic acid sequences of RT-PCR clones OSR8.401, OSR8.406, OSR8.403, OSR8.404, OSR8.402, OSR8,389 and OSR8.387;

Figure 3B shows the alignment of the non-coding regions of the preliminary nucleic acid sequences of RT-PCR clones OSR8.389, OSR8.387, OSR8.402 and OSR8.404;

Figure 4 shows the preliminary nucleic acid sequence of clones OSR8.401, OSR8.402 and OSR8.389;

Figure 5 shows the results of a northern blot of the class 2 clone, OSR8.402, comparing expression in the seed and during germination, using a random primed whole RT-PCR fragment as probe. In this Figure, L = leaf, C = cotyledons, S = seeds and B = buds;

Figure 6 shows the results of northern blots of the class 1 clone, CDCYS12 (using the coding region of the clone as a probe, labelled by PCR), the class 2 clone, CDCYS25 (using the non-coding region of the clone as a probe, labelled by PCR) and the class 6 clone, CDCYS66 (using the non-coding region of the clone as a probe, labelled by PCR) comparing expression in a range of developmental stages. In this Figure, B = bud, W = whole plant and L = leaf;

Figure 7 shows the alignment of the deduced amino acid sequences of clones OSR8.403, OSR8.404, OSR.402, OSR.389, OSR8.387, OSR8.406 and OSR8.401 with the published sequence of COT44 (designated CYS4.BRANA);

Figure 8 shows a schematic outline of the identification of cysteine protease isoforms using a cDNA library;

Figure 9 shows the alignment of the deduced amino acid sequences of the partial length cDNA class 2 clones, CYS2UP6.1, CYS2UP7.1 and CYS2UP8.2, with each other and with COT44 (designated CYS4.BRANA);

Figure 10 shows the alignment of the nucleic acid sequences of the partial cDNA clones, CYS2UP6, CYS2UP7, CYS2UP8, CYS6UP3, CYS6UP5, CYS6UP2 and CYS6UP4, with each other and with COT44;

Figure 11 shows the alignment of the full length cDNA clones, CDCYS66, CDCYS24, CDCYS22, CDCYS25, CDCYS12 and CDCYS14, with each other and with COT44;

Figure 12 shows the nucleic acid sequence of cDNA clone CDCYS12;

Figure 13 shows the nucleic acid sequence of cDNA clone CDCYS14;

Figure 14 shows the nucleic acid sequence of cDNA clone CDCYS22;

Figure 15 shows the nucleic acid sequence of cDNA clone CDCYS24;

Figure 16 shows the nucleic acid sequence of cDNA clone CDCYS25;

Figure 17 shows the nucleic acid sequence of cDNA clone CDCYS66;

Figure 18 shows the alignment of the predicted amino acid sequences of cDNA clones CDCYS12, CDCYS14, CDCYS22, CDCYS24, CDCYS25 and CDCYS66, and compares the sequences with the characterising features of plant cysteine proteases;

Figure 19 shows a promoter nucleic acid sequence from class 1 genomic clone pKS12p6.

Figure 20 shows a promoter nucleic acid sequence from class 2 genomic clone pKS25p7.

Figure 21 shows a promoter nucleic acid sequence from class 6 genomic clone PKS66P1.

Figure 22 shows the mapping of the transcription start by primer extension experiments as well as the position of the putative TATA box of clones pKS12P6 (class1), pKS25P7 (class2) and pKS66P1 (class 6). Numbers indicate the distance to the 5'end of the cDNAs.

Figure 23 shows the vector constructs for tobacco transformation.

Figure 24 shows the levels of GUS expression in shoot-generating calli during transformation.

Figure 25 shows the levels of GUS expression in leaves from primary transformants.

Figure 26 shows the time course of GUS expression in the progeny of 2 random primary transformants per class, 0 to 36 days after imbibition (DAI). NVS designates the wild type negative control.

Figure 27 shows the GUS activity in the progeny of 24 primary transformants of class 1. Seedlings were assessed at 0, 14 and 28 DAI. The Figure also includes the GUS activity in young leaves from the primary transformants.

Figure 28 shows the GUS activity in the progeny of 24 primary transformants of class 2. Seedlings were assessed at 0, 14 and 28 DAI. The Figure also includes the GUS activity in young leaves from the primary transformants.

Figure 29 shows the GUS activity in the progeny of 24 primary transformants of class 6. Seedlings were assessed at 0, 14 and 28 DAI. The Figure also includes the GUS activity in young leaves from the primary transformants.

In general, in the Figures the designations CO or COD in connection with a clone indicate the coding region, and NC or NCD indicates the non-coding region.

In outline, the Examples describe the amplification of a range of germination-expressed cysteine protease partial clones by reverse transcriptase polymerase chain reaction (RT-PCR) on germinating oilseed rape RNA. Preliminary assessment of the clones' expression in dry seeds and germinating seedlings by northern blotting, followed by more detailed northern blot experiments on selected clones to assess their time course of expression. A cDNA library was constructed from tissues showing a high expression of these clones, followed by screening of a genomic library to subclone the promoter areas. Final assessment of the spatial and temporal regulation of the cloned promoters was conducted by transcriptional fusion of the promoter fragments with the β-glucuronidase (GUS) reporter gene and transformation into tobacco.

Example 1 - CONSTRUCTION OF A CYSTEINE PROTEASE REVERSE TRANSCRIBED PCR LIBRARY FROM GERMINATING OILSEED RAPE SEEDLINGS

In order to identify germination specific sequences, a RT-PCR approach on oilseed rape seedling RNA was utilised. A reverse oligo(dT) primer and forward primers, designed to the cysteine protease coding regions conserved between several plant species, were used to amplify a 750 bp RT-PCR product covering 500 bp of the coding region and about 250 bp of the non-coding region. The 5'-end was sequenced to confirm the identity of the RT-PCR products as cysteine protease related clones. Since there is much less pressure of selection on non-coding regions, significant differences in the 3'-end non-coding regions may predict differences in the 3'-end untranslated regions with an effect on the promoter. This general approach is shown schematically in Figure 1.

Preparatory Methods :

Plant material

Five grams of oil-seed rape seeds (*Brassica napus*) from the variety Westar were sterilised in 1% sodium hypochlorite for 10 minutes. After several washes in sterile water, seeds were imbibed with sterile water for 12 hours at 4°C in the dark to synchronise the germination. They were sown on wet sterile Whattman paper and grown at 25°C in the dark prior to harvesting the cotyledons.

RNA extraction

Lithium chloride method

Total RNA was isolated from dry seeds and three days-old oil seed rape seedlings using a modification of the protocol described by Jepson *et al.*, (1991). Tissues (5 g) were ground with liquid nitrogen in a mortar and pestle until a fine powder was obtained. After addition of 9 ml homogenisation buffer [400 mM NaCl; 50 mM Tris-HCl, pH 9.0; 1 % SDS; 5 mM ethylene diamine tetraacetic acid (EDTA); 4 U/ml heparin; 1 mM aurintricarboxylic acid (ATT); 10 mM dithiothreitol (DTT)] and 4.0 ml phenol saturated in homogenisation buffer and supplemented with 10 % (v/v) m-cresol before use, the tissues were ground again until a fine paste was obtained. The paste was transferred to a cold corex tube and centrifuged for 15 minutes at 13,000 rpm (Sorval, SS34, 4°C). The supernatant was transferred to another tube, extracted for 5 minutes with 5 ml of phenol-chloroform and centrifuged for 30 minutes at 9,000 rpm (Sorval, SS34, 4°C) to recover the aqueous phase. After 3 phenol-chloroform extractions, the RNA was recovered by precipitating the supernatant overnight on ice with one fifth volume of 12 M lithium chloride. After centrifuging for 30 minutes at 9,000 rpm (Sorval, SS34, 4°C), the supernatant was removed and the pellet resuspended in 1 ml of 5 mM Tris (pH 7.5) prior to transfer to a microtube. After a second lithium chloride precipitation overnight, the pellet was washed twice with 70% ethanol, resuspended in 0.2 ml DEPC treated water and stored at -70°C.

Caesium chloride method

Total RNA was isolated from a range of developmental stages of oil seed rape seedlings using a protocol modified from Okayama *et al* (1979). Tissues (2-4 g) were ground in a mortar and pestle with 1 g Al_2O_3 in the presence of liquid nitrogen. The powder, kept in dry ice, was then mixed in a corex tube with 8 ml of pre-warmed (65°C) homogenisation buffer [5M thiocyanate guanidine; 0.5 %, w/v, lauryl sarcosine sodium; 0.025 M sodium citrate; pH 7.0] supplemented before use with 2.5 % (v/v) B-mercaptoethanol, and incubated at 40°C for 10 minutes with vortexing until complete defrosting of the tissues. After centrifugation at 12,000 rpm for 30 minutes at 15°C (Sorval, SS34 rotor), the supernatant was recovered, homogenisation buffer added to 8 ml and supplemented with 0.1 g CsCl per ml. After the CsCl had dissolved, the homogenate was added in a 12 ml polyalomer tube containing 2.5 ml of a CsCl high density cushion [5.7 M CsCl; 0.1 M EDTA] without disturbing the cushion. After centrifugation at 25,000 rpm for 24 hours at 20°C (Sorval, TH-641 rotor) the supernatant was removed by suction and the wall of the tube cleaned with absorbent paper before resuspension of the RNA loop into 300 µl resuspension buffer [7 M urea; 2%, w/v, lauryl sarcosin sodium]. The RNA was then transferred to a 1.5 ml tube and extracted with an equal volume of phenol and an equal volume of chloroform/isoamylic alcohol [24:1, v/v]. Following centrifugation at 13,000 rpm for 5 minutes the aqueous phase was recovered and extracted again with an equal volume of chloroform. The RNA was precipitated overnight at -20°C by adding a one-tenth volume of 3 M sodium acetate and 2.5 volumes of cold ethanol. After centrifugation at 13,000 rpm for 15 minutes at 4°C, the supernatant was discarded and the pellet was washed with 1.5 ml of cold 70% ethanol. The centrifugation was repeated for 10 minutes, the supernatant was discarded and the pellet dried for 3-4 minutes in a speedvacuum. The pellet was resuspended in sterile DEPC-treated water, the RNA precipitated again and the pellet washed, as described above, and finally resuspended in 50 µl DEPC-treated water and stored at -70°C.

Labelling of DNA probes

Terminal exchange

The oligos (25-50 ng) were labelled by phosphorylating their hydroxylated 5'-end using 20 U T4 polynucleotide kinase (New England Biolabs) in a 25 µl reaction [30 uCi [γ-^{32}P] ATP (Amersham, 5000 Ci/mmol), 1X kinasing buffer] at 37°C for 30 minutes. The oligos were purified from the unincorporated nucleotides using G-25 sephadex spun columns (5prime→3prime, Inc®), as recommended by the manufacturer.

PCR

Plasmid DNA (5 ng) was amplified [(94°C, 1 minute; 65°C, 1 minute; 72°C, 1 minute) x 17 cycles; (72°C, 7 minutes) x 1 cycle] using 2.5 U of Taq polymerase (Gibco BRL) in a 50 µl PCR reaction [0.25 µM [α-^{32}P] dATP(Amersham, 3000 Ci/mmol); 0.4 µM dATP; 50 µM other dNTPs; 1.5 mM MgCl2; 0.5 µM oligos; 1X PCR buffer]. The probes were purified from the unincorporated nucleotides using G-50 sephadex spun columns (Pharmacia Biotech), as recommended by the manufacturer.

Random priming

Plasmid DNA (25-50 ng) was labelled using the "Oligolabelling Kit" (Pharmacia) following manufacturer's recommendations.

Northern blot hybridization

Northern blot experiments were performed according to Sambrook *et al.* Total RNA (10 µg) was mixed with 2.5 vol of loading buffer (5prime→3prime, Inc®), sized by electrophoresis on a 1.2 % agarose denaturing gel [1X MOPS(40 mM MOPS, pH 7.0; 10 mM sodium acetate; 1 mM EDTA); 17%, v/v, formaldehyde] in alkaline running buffer [1X MOPS; 7%, v/v, formaldehyde] and transferred to nylon membranes (Hybond N, Amersham) by capillary blotting according to the manufacturer's recommendations. Hybridizations were performed overnight at 65°C in hybridisation buffer [5XSSPE (900 mM NaCl; 50 mM $NaH_2PO_4.H_20$; 5 mM EDTA; pH 7.4, NaOH); 0.5% sodium dodecyl sulphate (SDS); 1% powdered milk] supplemented with 100 µg/ml denatured salmon sperm DNA. Blots were

washed in 3X SSC [20X SSC: 450 mM NaCl; 45 mM Na_3citrate.$2H_2O$; pH 7.0, HCl] and 1% SDS for 1 hour at 65°C, and exposed to X-ray films (X-OMAT AR, Kodak) at -80°C with intensifying screen.

RT-PCR Library Construction and Screening

A RT-PCR library was constructed from germinating oilseed rape seedlings using a "Perkin Elmer GeneAmp® RNA PCR Kit" and following the manufacturer's recommendations. Total RNA (1 µg) extracted from cotyledons 3 days after seeds imbibition was reverse transcribed [65°C, 2 minutes; 42°C, 30 minutes; 99°C, 5 minutes; 6°C, 5 minutes] using an oligo(dT) primer (MPRACE1B) stabilised for the PCR. The heteroduplex DNA-RNA was subsequently amplified by PCR using the same downstream primer as for the reverse transcription and a set of two upstream degenerate primers (DEGCYS1 and DEGCYS2) designed on the peptide level from a motif of cysteine protease coding regions conserved among most of the plant species [GCNCCLM(NED)]. Cycling conditions: [(95°C, 2 minutes; 55°C, 2 minutes; 72°C, 1.5 minutes) x 2 cycles; (95°C, 2 minutes; 55°C, 1 minute; 72°C, 1.5 minutes) x 33 cycles] followed by 7 minutes at 72°C.

The RT-PCR products were ligated into a pCRII vector before transforming *E. coli* using the "TA Cloning® Kit" (Invitrogen) following the manufacturer's recommendations. This system takes advantage of the non-template dependent activity of thermostable polymerases used in PCR that add a single deoxyadenosine to the 3'-ends of all duplex molecules provided by PCR. This allows direct cloning into a pCRII vector which contains overhanging deoxythimidine. Plasmid DNA was purified using "Wizard DNA miniprep DNA purification system"(Promega) and sequenced by the chain termination method (Sanger *et al.*, 1977) with "Sequenase version 2.0 T7 DNA polymerase" (USB) according to the manufacturer's recommendations.

OLIGOS	**SEQUENCE (5'→ 3')**
MPRACE1B	GGC CAC GCG TCG ACT AGT TAC TCG AGT TTT TTT TTT TTT TTT T
DEGCYS1	GGI TG(CT) AA(CT) GGI GGI (CT)TI ATG
DEGCYS2	GGI TG(CT) AA(CT) GGI GGI (CT)TI ATG (GA)A

Analysis

The RT-PCR products were analysed by electrophoresis on an agarose gel. As can be seen in Figure 2, two major 750 bp fragments of the expected size, amplified from cotyledons RNA, but not from seeds RNA were found. These were gel excised and cloned. In a parallel approach, aliquots of the RT-PCR products without gel purification were cloned to generate a three days-old expressed cysteine protease library containing 400 clones from which 22 came from gel excision. A colony screening with the oligos used for the RT-PCR identified 250 putative cysteine proteases. Seven clones from gel excision were taken at random and fully sequenced, all of them were cysteine proteases, and fell into 3 classes.

These clones were given the following designations :

CLONES	CLASS
OSR8.401	1
OSR8.402	2
OSR8.403	2
OSR8.404	2
OSR8.406	6
OSR.387	6
OSR.389	6

The preliminary DNA sequences of these clones and their alignment is shown in Figures 3A and 3B.

Three clones, one per class, OSR8.401 (class 1); OSR8.402 (class 2) and OSR8.389 (class 6) whose preliminary DNA sequences are shown in Figure 4, were labelled by random priming and assessed on northern blots containing total RNA extracted from seeds and cotyledons using the lithium chloride method. They appeared to be well expressed during germination but not in dry seeds. Furthermore, some cross-hybridisation was observed

between the different classes due to the nature of the probe. This is illustrated in Figure 5 which shows the result for the class 2 clone, OSR8.402, with a random primed whole RT-PCR fragment as probe.

The library was then screened using the five cysteine protease clones labelled by random priming. No new classes were clearly identified, all the diversity of clones was present in the gel-excised fragments. The expression pattern of the three classes of clones was assessed by northern blots with RNA, extracted with the caesium chloride method, from a range of developmental stages as shown in Figure 6. For class 2 and 6, hybridisations were performed using the 3'-non coding regions of clones OSR8.402 and OSR8.389 respectively, to avoid any cross-hybridisation. For class 1, a part of the coding region of OSR8.401 clone was used as a probe because no non-coding region was available in the RT-PCR clones. To increase the specificity, probes were labelled by PCR using the oligo(dT) reverse primer (MPRACE1B) and an internal reverse primer (CYS8.402 and CYS8.389) for classes 2 and 6, whilst using 2 internal primers (CYS8.401 and CYS8.401R) for class 1. Classes 2 and 6 are expressed following seed imbibition and for the first 4-5 days of early seedling growth but are not expressed in mature plant organs or in the developing seed. Class 1 shows some expression in buds and leaves but this may be the result of some cross-hybridisation due to the nature of the probe. Class 2 is highly related to COT44, the only cysteine protease published for oilseed rape (Comai and Harrada, 1989). Figure 7 shows alignment of deduced amino acid sequences of the clones with COT44.

OLIGOS	SEQUENCE (5'→ 3')
CYS8.401	TAT CCT TAT CAA GAA CGT GAT GGC A
CYS8.401R	CCT ACG ATG AGC ACT GCG TGG T
CYS8.402	GCA GTA ATC AAA TTG GGA TTG TTA TAA
CYS8.389	CGT GGA ACC AGC AGT GTT TGA AGT T

Example 2 - CONSTRUCTION OF A STANDARD OLIGOd(T) PRIMED AND A CYSTEINE PROTEASE SPECIFICALLY PRIMED cDNA LIBRARY FROM GERMINATING OILSEED RAPE SEEDLINGS

As the RT-PCR products were not full length and to avoid PCR generated mutations, an oilseed rape cDNA library was constructed from a developmental stage showing a high expression of the three clones of interest. A specifically primed library was constructed using two specific oligos designed on the basis of the three classes of RT-PCR clones, rather than using an oligo(dT) primer. As a result only the cysteine protease clones were reverse transcribed, which provided a small number of short clones of about 650 bp, all of them being full length at the 5'-end. This allowed the design of oligos to the 5'-non-coding regions for use in screening a standard oligo(dT) primed cDNA library directly for full length clones. This general approach is shown in schematically in Figure 8.

Preparatory Methods :

Plant material

Five grams of oilseed rape seeds (*Brassica napus*) from the variety Westar were sterilised in 1% sodium hypochlorite for 10 minutes. After several washes in sterile water, seeds were imbibed with sterile water for 12 hours at 4°C in the dark to synchronise the germination. They were sown on wet sterile Whattman paper and grown at 25°C in the dark for 2 days prior to harvesting the cotyledons.

RNA extraction and purification

Total RNA was isolated from 2 days-old oil seed rape seedlings using the caesium chloride method described previously. Polyadenylated RNA was purified from 1 mg of total RNA using a "PolyATract mRNA isolation system" (Promega), according to the manufacturer's recommendations. The system uses a biotinylated oligo(dT) primer to hybridise at high efficiency in solution to the 3'-poly(A) region of the mRNAs. The hybrids were then captured and washed at high stringency using streptavidin coupled to paramagnetic particles and a magnetic separation strand, prior to elution with water.

cDNA Library Construction and Screening

The standard oligo(dT) primed cDNA library and the cysteine protease specifically primed library were constructed from 2 days-old oilseed rape poly(A) RNA (5 μg), using a lambda "ZAP-cDNA® Synthesis Kit" (Stratagene). The manufacturer's recommendation were followed strictly although, for the specific library, the reverse transcription was primed using a mix of two specific oligos (CDNA8.401R and CDNA8.387R) respectively for the class 1 and for classes 2 and 6 of the RT-PCR clones, modified to include a *Xho* I site at their 5'-end. The second strand was synthesised by nick-translation using DNA polymerase I, after treatment of the heteroduplex with RNase H. The cDNAs were filled in with Klenow, ligated to *EcoR* I adapters and digested with *Xho* I prior to size-fractionation on Sepharose®-400 spun column (Pharmacia) and directional cloning as an *EcoR* I-*Xho* I insert into the polylinker of pBluescript phagemid contained within the Uni-ZAP vector arms. Lambda-ZAP is a replacement lambda which has been engineered to contain pBluescript phagemid, which polylinker is used for cloning the cDNAs.

OLIGOS	SEQUENCE (5'→ 3')
CDNA8.401R	GAG AGA GAG AGA GAG AGA GAA CTA GTC TCG AGT CCC ATG GTT TTT AAT
CDNA8.387R	GAG AGA GAG AGA GAG AGA GAA CTA GTC TCG AGC CGC CGT TTT TCA T

The library was packaged *in vitro* using Gigapack® II Gold packaging extract (Stratagene), plated on *E. coli* cell line XL1-Blue MRF' and transferred onto nylon membranes (Hybond N, Amersham) according to the manufacturer's recommendations. Labelling of probes, hybridisations and washes were performed as described previously.

Selected lambda-ZAP clones were excised *in vivo* to recover the cloned cDNA as a phagemid in pBluescript SK⁻. *E. coli* SolR cells were co-transfected with the recombinant lambda-ZAP and with a helper phage which provided the proteins necessary for the synthesis of a single strand of DNA which, once circularised, provide a functional phagemid.

Analysis

a) A small cysteine protease-enriched specifically primed cDNA library containing 1.10^4 plaque forming units (pfu) was obtained from 2 days-old oilseed rape cotyledons. 5.10^3 pfu were plated and transferred onto three replicate membranes to check for cross-hybridisation. Membranes were screened with 3 oligos (CYS8.401MR, CYS8.402MR, CYS8.406MR) designed respectively to the 5'-end of classes 1, 2 and 6 of the RT-PCR clones and labelled as described above.

Five similar but distinct 5'-end cysteine protease cDNAs containing a short 5'-untranslated region and 650 bp of coding region were obtained, excised *in vivo* and sequenced. They fall into class 2 and class 6 but no 5'-end cDNA was found for class 1. Although clones from class 2 are highly related to COT44 cDNA (Comai and Harrada, 1989), their 5'-end is 160 bp longer. This indicates that COT44 is missing, 46 amino acids (aa) corresponding to the signal peptide and to a part of the propeptide. This is further illustrated by Figure 9 which shows the alignment of the deduced amino acid sequence of the cDNA clones CYS2UP6, CYS2UP7 and CYS2UP8 from class 2 with COT44.

Two oligos (CYS6B-UP and CYS6A-UP) were designed to the 5'-non-coding region of respectively, cDNAs from classes 2 and 6, to screen the standard oligo(dT) primed library directly for full length clones. The preliminary sequence alignment of the clones with each other and COT 44 is shown in Figure 10.

OLIGOS	SEQUENCE (5'→ 3')
CYS8.401MR	CCC ATG GTT TTT AAT GAC AAA TTG AAA A
CYS8.402MR	CCG CCG TTT TTC ATT ATG AAT TGA AA
CYS8.406MR	CGC CGT TTT TCA TGA TGA ATT GAA AA

b) A representative oligo(dT) primed cDNA library containing 1.10^7 pfu was obtained from 2 days-old oilseed rape cotyledons. The size of the inserts, estimated by PCR, is ranging from 0.75 kb to 3 kb with an average insert size of 1.5 kb. 1.10^6 pfu were plated and transferred onto three replicate membranes to check for cross-hybridisation. Membranes were first screened with a 5'-end oligo (CYS8.401MR, CYS6B-UP and CYS6A-UP) designed respectively to the 5'-end coding region of class 1 RT-PCR clones and to the 5'-

untranslated region of class 2 and class 6 cDNAs to identify full length clones for class 2 and 6 (class 1, 20 positives; class 2 , 250 positives; class 6, 130 positives).

OLIGOS	SEQUENCE (5'→ 3')
CYS8.401MR	CCC ATG GTT TTT AAT GAC AAA TTG AAA A
CYS6B-UP	TAG AAA ACC AAC AAA ACA AAC ATA CAA T
CYS6A-UP	GAA CAA CCA AGC CAA ACA TAC AAT AT

The same membranes were screened again with the 3'-end probes used on the developmental northern blots to ensure the correct form was chosen. For classes 2 and 6, hybridisations were performed using the 3'-non-coding regions labelled by PCR to avoid any cross hybridisation. For class 1, a part of the coding region was used as a probe because non-coding region was not available in the RT-PCR clone (class 1, 84 positives; class 2, 205 strong positives; class 6, 85 strong positives).

Comparison between the number of clones identified with both specific probes (5' and 3') confirmed that most of the cysteine protease cDNA clones present in the library are full length.

Ten clones per class which hybridised with (5' and 3' probes) but did not cross hybridise with probes from the two other classes were plaque purified and four clones per class were excised *in vivo*. Six full length cysteine protease cDNA clones falling into the three classes of cysteine proteases identified from the RT-PCR work were isolated and fully sequenced (two for class 1, three for class 2, and one for class 6). The alignment of the cDNA clones is shown in Figure 11. They fall into 3 classes of CP related to the papain super family and the pre-proenzymes share 52% (class1), 90% (class 6) and 96% identity (class 2) with cot44. Figures 12 to 17 show the nucleic acid sequences of clones CDCYS12, CDCYS14, CDCYS22, CDCYS24, CDCYS25 and CDCYS66 respectively. The peptide sequences were predicted and showed the characteristic features present in most of the plant cysteine proteases, as shown in Figure 18 for clones CDCYS12, CDCYS14, CDCYS22, CDCYS24, CDCYS25 and CDCYS66.

Example 3 - SCREENING OF AN OILSEED RAPE GENOMIC LIBRARY AND SUBCLONING AND CHARACTERISATION OF THE PROMOTER REGIONS

Oligonucleotide probes were generated to the 5'-end non-coding region of one cDNA clone per class and used to screen a genomic library in order to isolate clones carrying the promoter regions. For each class, genomic clones were isolated and the promoter subcloned into a phagemid for more precise characterisation and deletion.

Genomic Library Construction and Screening

An amplified λEMBL-3 random genomic library (Clontech) from oilseed rape (*Brassica napus* cv. Bridger) was constructed. DNA was partially digested with *Mbo* I and the fragments were separated on a sucrose gradient to produce size range between 8 to 22 kb before cloning into the *Bam*H I site of a λEMBL-3 replacement vector. The library was plated on *E. coli* strain LE392 cells. Screening and plaque purification were performed as described by Sambrook *et al.* (1989). Genomic clones corresponding to the three classes of cDNAs were isolated and λDNA minipreparations were carried out using a protocol from Grossberger (1987). Genomic clones were mapped using their restriction fragment length polymorphism (RFLP) patterns: clones were digested with a set of restriction enzymes, analysed on a 0.8% agarose gel and simultaneously transferred onto two membranes (Hybond N, Amersham) according to Sambrook *et al.*, (1989) prior to hybridisation.

Analysis

For the primary screening twenty genome-equivalent (2.10^6 pfu) were plated and transferred onto three replicate membranes to check for cross-hybridisation. Membranes were hybridised with 3 oligos (CDNA12, CDNA25 and CDNA66) designed respectively to the 5'-end of classes 1, 2 and 6 of the cDNA clones, to get as close as possible to the promoter area (class 1, sixteen strong positives; class 2, nine strong positives; class 6, eight strong positives). No cross-hybridisation was detected between the three classes and ten clones per class were chosen for a secondary screening.

OLIGOS	SEQUENCE (5'→ 3')
CDNA12	ATC GTC TTC TTC CTT TGT TTC TCT CA
CDNA25	CTT CGT CAG CGA AAC TCC TCT CTT
CDNA66	CAG AAC TAG AAC AAC CAA GCC AAA C

The secondary screening was performed using 3'-end PCR probes as described in connection with Figure 6, to specifically detect the genomic clones corresponding to the RT-PCR clones assessed by northern blots. Some of the clones were not identified by these probes (one for class 1 and two for class 6), since these were likely to be short clones in their 3'-end, thus useful for promoter isolation, they were rescued using the probes from the primary screening. Ten clones per class were plaque purified by two more rounds of purification using the probes from the primary screening. In order to avoid redundant clones (amplified, random library), DNA was prepared from eight genomic clones per class and characterised by RFLP. Clones were cut with *Sal* I and *Bam*H I, analysed on a 0.8% agarose gel, transferred onto replicate membranes and hybridised with two sets of probes. Oligos were designed to the 5'-non-coding region (CDNA12, CYS6B-UP and CYS6A-UP) and to the middle of the coding region (CYS8.401MR, CYS8.406MR and CYS8.406MR) of cDNA clones CDCYS12, CDCYS25 and CDCYS66 respectively.

Four remaining clones per class (12g4, 12g5, 12g6, 12g8; 25g2, 25g4, 25g5, 25g7; 66g1, 66g4, 66g8 and 66g9) were further characterised using another round of digestion/hybridisation. Class 1 cDNAs contain a *Bgl* II site 500 bp from the translation start and class 2 and 3 cDNAs contain a *Hind* III site 300 bp from the translation start, these enzymes were used in association with and without *Sal* I, which releases the insert, to generate genomic fragments suitable for subcloning. PCR experiments were carried out on genomic DNA and cDNAs to predict the size of the promoter area by identifying putative introns. PCR, using a forward primer in the 5'-non-coding region and a reverse primer located after the *Bgl* II and *Hind* III restriction sites, showed the presence of a 400 bp intron within the first 600 bp of class 1 cDNAs whilst no intron was present within the first 300 bp of class 2 and class 6. Promoter fragments with a predicted size in the range of 2-5 kb were identified for one genomic clone per class (12g6, 25g7 and 66g1), ready to be subcloned into pBluescript KS^+.

Example 4 - CHARACTERISATION OF TRANSCRIPTION STARTS

Genomic lambda-fragments containing the promoter were subcloned into pBluescript KS$^+$ for more precise characterisation. Sequencing allowed the identification of putative transcription signals before mapping the actual transcription start by primer extension experiments. This involved the extension of a labelled reverse primer designed in an area close to the translation start. After degradation of the RNA template the extension products were sized in a polyacrylamide gel.

Analysis

Genomic fragments containing the promoter were subcloned into pBluescript as a *Bgl* II-2.6 kb insert cloned in *Bam*H I for class 1 (pKS12P6), as a *Hind* III-4.2 kb insert for class 2 (pKS25P7) and as a *Bam*H I-*Hind* III-2.4 kb insert for class 6 (pKS66P1). Sequencing with pUC1 and pUC4 vector oligos and with two internal reverse primers designed to the 5'-end of the cDNAs (CDNA14R for class 1 and CDNA66R for class 2 and 6), allowed the orientation of the clones and the identification of putative transcription signals (Pautot *et al.*,1989). The full nucleotide sequence of the promoters from the sub-cloned genomic fragments is given in Figures 19, 20 and 21 respectively for class 1, 2 and 6.

OLIGOS	SEQUENCE (5'→ 3')
CDNA14R	GAA GAA ACT AGA AGA AGG GAG AAG AA
CDNA66R	TCA CTT CTT CAT CGG TTC TCC A

Transcription starts have been mapped precisely by primer extension experiments according to Sambrook *et al.* (1989) modified as follows. Oligos (CYSGE12R, CYSGE25R and CYSGE66R) designed respectively for class 1, 2 and 6 were labelled by terminal exchange as described previously. Total RNA (50 ug), isolated from 3 days-old oilseed rape cotyledons using the caesium chloride method, was precipitated together with 2 ng of primers and resuspended in 30 µl hybridisation buffer [1 mM EDTA; 400 mM NaCl; 40 mM Pipes, pH 6.4; 70%, v/v, deionised formamide]. Annealing was performed overnight at 32°C

following denaturing at 85°C for 10 minutes. After precipitation and resuspension in 25 µl of reverse transcription buffer [50 mM KCl; 10 mM $MgCl_2$; 1 mM dNTPs; 1 U/ul RNase inhibitor], the primers were extended for 90 minutes at 42°C with 2.5 U MuLV reverse transcriptase (Perkin-Elmer). Template RNA was degraded for 30 minutes at 37°C with 20 U of RNase (RNace-it, Stratagene). For each class, the extension products were analysed on a polyacrylamide denaturing gel in parallel with a sequencing reaction performed on the genomic clones (pKS12P6, pKS25P7 and pKS66P1) using the same primers as for the primer extension.

OLIGOS	SEQUENCE (5'→ 3')
CYSGE12R	AGG AAG AAG ACG ATG ATG GTG ACA
CYSGE25R	GTA CAA GAG AAG TAA AGA GAG GAG T
CYSGE66R	CGT ATA GGA GAA GTA AAG AAA TGA GT

As shown in Figure 22, class 1 transcription start is in a good context [$t_{27}T_{35}C_{49}\underline{A}_{78}a_{18}C_{45}g_{8\%}$] compared to 49 other plant genes compiled by Pautot *et al.*(1989). Conserved nucleotides are in capitals, important ones are in bold and the transcription start point is underlined. This transcription start has been mapped 22 nucleotides downstream a putative TATA box localised 179 nucleotides before the translation start and 152 nucleotides upstream the longest cDNA.

Although the consensus for the TATA box is not optimal [$c_{34}a_{18}t_{32}t_{34}\mathbf{a_3}A_{97}T_{90}A_{94}a_{47}A_{95}a_{30}A_{71}G_{44\%}$] compared to 79 other plant genes compiled by Pautot *et al.* (1989), this result is a confirmation of a previous primer extension experiment using an oligo priming 34 nucleotides downstream the ATG. The distance between the transcription start and the longest cDNA might be explained by the presence of an intron within the 5'untranslated region or by the existence of an alternative transcription start point.

The Class 2 transcription start is in a good context [$a_{18}a_{20}\mathbf{a_{22}}\underline{A}_{78}T_{49}C_{45}A_{43\%}$], and has been localised 33 nucleotides after a putative TATA box fitting very well within the plant consensus [$T_{37}g_{11}g_{14}t_{34}T_{96}A_{97}T_{90}A_{94}a_{47}A_{95}T_{63}A_{71}G_{44\%}$]. This corresponds to 53 nucleotides before the translation start and 26 nucleotides upstream the cDNA (Figure 22).

Class 6 transcription start is nearly in the same context as class 2 [$g_{18}a_{20}\mathbf{a_{22}}\underline{A}_{78}T_{49}C_{45}A_{43\%}$] and has been localised 30 nucleotides after a putative TATA box

showing exactly the same consensus as for class 2. This corresponds to 51 nucleotides before the translation start and 19 nucleotides upstream the cDNA (Figure 22).

Example 5 - PROMOTER EXCISION

Prior to fusion with the reporter genes, promoters must be cut precisely between the transcription start and the translation start. Since no useful restriction site was available for the class 2 and 6 genomic clones, a site was engineered into a PCR fragment used to replace a corresponding endogenous fragment.

Analysis

A *Hind* III site was introduced by PCR on class 1 genomic clone, 2 nucleotides before the translation start, in order to eliminate the remaining part of coding region. The 270 bp fragment was generated by 15 cycles of PCR on pKS12p6 DNA, using CYSGE12C and CYSG12CR oligos.

In the same way, a *Bam*H I site was introduced before the translation start of class 2 and class 6 genomic clones using 2 sets of oligos (CYSGE25C, CYSG25CR and CYSGE66C, CYSG66CR) respectively on pKS25p7 and pKS66p1 DNA to generate a 165 bp fragment.

For class 1, pKS12p6 was cut with *Bsm* I and *Hind* III and gel recovered to excise a 900 bp fragment prior to replacement with the PCR fragment cut by *Bsm* I and *Hind* III, to generate pKS12P.

For class 2, pKS25p7 was cut with *Sph* I and *Bam*H I, gel recovered to excise a 460 bp fragment and ligated to the replacement PCR fragment cut by *Sph* I and *Bam*H I, to generate pKS25P.

For class 6, pKS66p1 was cut with *Hind* III, filled in with Klenow fragment of DNA polymerase I, cut with *Sph* I and gel recovered to excise a 440 bpfragment. The PCR fragment was blunt ended with T4 DNA polymerase, cut with *Sph* I and cloned into the deleted pKS66p1 to generate pKS66P.

OLIGOS	SEQUENCE (5'→ 3')
CYSGE12C	GTA ATG GCC TAG CCT GTC TGG C
CYSG12CR	GAT GAT GGT GAC AAG CTT TTT CTT ACA GG
CYSGE25C	CTA TCT TGC ATG CCC ATT ATT ACT TT
CYSG25CR	ACG AAG CCG GAT CCT ATG TTT GTT TTG TTG
CYSGE66C	CAT CTT GCA TGC CCA TTA CTG CAT
CYSG66CR	AGG AAG CCG GAT CCT ATG TTT GGC TTG G

Example 6 - PROMOTER-GUS CONSTRUCTS

In order to assess the spatial and temporal regulation of the cloned promoter regions in an heterologous system and under different biotic and environmental conditions, they were used them to drive a reporter gene into tobacco. Transcriptional fusions between each promoter fragment and the β-glucuronidase (GUS) gene were engineered for plant assays and histochemical localisation. For the avoidance of doubt, a reporter gene is used here for convenience only, and to demonstrate the principles involved. In non-test situations the gene controlled by the promoter of the present invention will be that which produces the desired effect.

Plasmid Construction

Standard recombinant DNA methods were adopted in the construction of plasmid vectors (Sambrook *et al.*, 1989). The CP12 *Hind* III-*Not* I-1.7 kb promoter fragment was excised from pKS12P, filled in using Klenow fragment of DNA polymerase I and ligated into the *Sma* I site of the *Agrobacterium* Ti vector pTAK1 containing the *E.coli uidA* gene encoding β-glucuronidase (Jefferson *et al.*, 1987), to produce a pTAKCP12 binary vector. In the same way CP25 *Hind* III-*Bam*H I-3.7 kb and CP66 *Bam*H I-1.9 kb promoter fragments were excised from pKS25P and pKS66P respectively, and ligated into pTAK1 cut with the same enzymes to produce pTAKCP25 and pTAKCP66 binary vectors. All constructs were transformed into *E. Coli* strain DH5α as an intermediate host for the vectors construction.The structure of the resultant chimeric reporter gene constructs was verified by PCR, restriction digest and sequence analysis. Figure 23 shows a schematic of the constructs for plant transformation.

Plant Transformation

Plasmids pTAKCP12, pTAKCP25 and pTAKCP66 were transferred into *Agrobacteriun tumefaciens* LBA4404 using the freeze/thaw method described by Holsters *et al.* (1978). Tobacco (*Nicotiana tabacum* var. Samsun) transformants were produced by the leaf disc method (Bevan, 1984). Shoots were regenerated on medium containing 100mg/l kanamycin and 200 mg/l carbenicillin. After rooting, plantlets were transferred to the glasshouse and grown under 16 h light/ 8 h dark conditions.

Results

	12CP plants	25CP plants	66CP plants
# shoot taken	440	250	404
# shoot rooting	130 (30%)	82 (33%)	120 (30%)
# shoot subdivided	100	82	100
# shoot re-rooting	84 (84%)	58 (70%)	71 (71%)

PRIMARY TRANSFORMANT ANALYSIS

Objective

Primary transformants were first analysed by polymerase chain reaction (PCR) in order to reduce the number of plants to analyse and to make sure they contained the intact promoter-reporter gene cassette.

Since promoters can be deregulated in callus, GUS analysis were carried out on calli from transformation to make sure the promoters are not active at this stageas this might have had a deleterious effect on future transformations efficiency depending on the nature of the transgene (e.g. if barnase ribonuclease gene is driven by one of these promoters). Promoter activity was also assessed in young leaves from primary transformants to confirm the absence of ectopic expression at this stage.

Polymerase chain reaction

Genomic DNA for PCR analysis of transgenic plants was prepared according to Edwards *et al.* (1992). Plant extracts DNA (2.5 ul) was amplified [hot start at 80°C ;(94°C, 1

min; 63°C, 1 min; 72°C, 1 min) x 35 cycles; (72°C, 7 min) x 1 cycle] using 2 U of Taq polymerase (Gibco BRL) in a 25 ul PCR reaction [200 uM dNTPs; 3 mM MgCl2; 1 uM oligos; 1X PCR buffer].

Results

A total of 37 individual transformants per class were randomly picked from *in-vitro* culture 12CP, 25CP and 66CP explants and analysed with 2 sets of primers. The first set contained one primer specific to the 5' end of the NOS terminator of the NPTII gene (NOSTER1) and a reverse-primer specific to the 5'end of the cloned promoters (CYSGE12R, CYSGE25R or CYSGE66R). The second set contained one primer specific to the 3'-end of the promoters (CYSGE12C, CYSGE25C or CYSGE66C) and a reverse-primer specific to the 5' portion of the GUS gene (GUS1R). A total of 34 explants was found to be double PCR positive for class 6 (94% of the plants tested), while for class 1 and class 2 only 27 plants gave the expected result (73% of the plants tested). Plants containing the intact cassette were transferred to the glasshouse and self-pollinated.

OLIGOS	SEQUENCE (5'→ 3')
CYSGE12RT	GGG TTC TTC TGG GTA GCA AAC TG
CYSGE25RT	ACT TCA CGT TCT GAA TCT CAT CGA A
CYSGE66RT	GGG CCA GAA TGC GGA TTT TAC TAA
GUS1R	CGC TTT CCC ACC AAC GCT GAT C
NOSTER1	TTG AAT CCT GTT GCC GGT CTT GC

GUS enzyme assays

Fluorometric assays for GUS activity performed with the substrate 4-methylumbelliferyl-D-glucuronide (Sigma) were carried out using a Perkin-Elmer LS-35 fluorometer (Jefferson *et al.*, 1987). Protein concentration of tissue homogenates were determined by the Bio-Rad protein assay (Bradford, 1976).

Results

GUS assays were carried out for each class on 20 regenerating calli resulting from the transformation process. Tobacco extracts from wild type leaves and callus as well as from leaves from 35S-GUS transgenic plants were used respectively as negative and positive

controls. As shown in Figure 24, no significant GUS activity could be detected in calli compared to the levels presents in leaves from the 35S-GUS plants.

Figure 25 shows a preliminary assessment of the levels of GUS activity in young leaves from primary transformants of each class. Results indicate that the promoters are not active at this stage. This is a confirmation of the northern blot results obtained for class 2 and 6 but contradicts those for class 1 (Figure 6). Class 1 CP showed some expression in leaves but it is thought to be due to a cross-hybridisation problem due to the nature of the probe. The GUS result seemed to confirm the latest hypothesis.

ANALYSIS OF SEGREGATING POPULATIONS

Objective

The objective was to select for each class, four GUS expressing lines ranging from low expressors to high expressors, preferably from lines with a single locus insertion of the transgene as this facilitates the comparisons between lines. The number of loci in the primary transformants is estimated by the segregation of the NPTII (kanamycin-resistance) gene in the progeny.

Segregation test

Seeds were sterilised in 10% bleach for 15 min. After several washes in sterile water around 150 seeds were sown on ½ MS medium (2.3 g/l MS salt, 1.5% sucrose, 0.8% Bactoagar, pH 5.9) containing 100 mg/l kanamycin. Seeds were grown.for three weeks at 26^{0}C with 16 hours/8 hours light/dark prior to scoring. If the primary transformants contained one copy of the transgene, the expected ratio for kan^R to kan^S seeds was 3 to 1 (although in very rare cases one locus could possibly contain several transgenes).

Results

A substantial number of plants are showing a *petaloïdie* phenotype in which a variable proportion of the flowers in a plant have one or more normal stamens replaced by petals. Some of these plants were so badly affected that we could not recover any seeds.

The table below summarises the genetic data for the primary transformants.

	12CP plants	25CP plants	66CP plants
# lines in glasshouse	27	27	34
# lines giving seeds	22 (81%)	22 (81%)	29 (85%)
# *petaloïdie* phenotype	13 (48%)	11 (41%)	5 (15%)
# single loci insertion	11 (50%)	9 (41%)	12 (41%)

Preliminary time course experiment

The northern blot results indicated an accumulation of CP mRNA in oil seed rape at 2 to 3 days after seeds imbibition (DAI). However, the heterologous expression in tobacco may differ from the endogenous expression in oil seed rape due to differences in physiology and transcription machinery. Furthermore, the activity of the promoter was indirectly analysed through a reporter protein, which delays the detection. So, in order to work out at which time point all the F1 generations should be analysed, the time course of GUS expression from each promoter had to be established.

Results

The experiment was carried out on 2 random lines per class as well as on wild type and 35S-GUS control lines. For each time point, 40 seeds of lines 12CP5, 12CP14, 25CP8, 25CP13, 66CP8 and 66CP76, as well as the controls, were grown at 26^{0}C with 16 hours/8 hours light/dark, on plates containing 1/2 MS media. Seedlings were sampled at 0, 2, 4, 6, 8, 10, 29 and 36 DAI and stored at -70°C. At time point 0 to 10 DAI, the 10 biggest seedlings were collected while only 5 were taken at 29 and 36 DAI which increased the variability within these samples.

In tobacco seedlings grown under the conditions described above, three lines out of 6 showed an induction of GUS expression during germination (Figure 26). The activity peaked at around 30 DAI, although the protein accumulation clearly started at 8 DAI. Thus, the F1 generations should be assessed at 14 and 28 DAI to identify a range of GUS expressors. The maximum of activity was about 5% of 35S-GUS in the best expressing line (12CP5) although the relatively high level of expression in leaves suggests that this might be due to a position

effect of the transgene. Normal levels of expression are more likely to be around 1% of 35S-GUS.

Identification of high GUS-expressing lines during germination.

Results

Seedlings were grown on 1/2 MS media supplemented with 100 mg/l kanamycin in the conditions described previously. Five seedlings were harvested at 0 (dry seeds), 14 (2 expanded leaves) and 28 DAI (4 expanded leaves), pooled and assessed in duplicates as described previously.

Figures 27, 28, and 29 summarise the expression levels for class 1, 2 and 6 respectively. These preliminary data suggest that the promoters are expressed in a seedling-specific manner in tobacco. As expected from the RNA study in oilseed rape the levels of expression are low. Class 2 promoter fragment is more active than class 1 at this stage, while class 6 gives extremely low levels of expression.

GUS histochemical detection

GUS histochemical staining of whole seedlings was achieved by vacuum infiltration with a solution containing 1 mM X-gluc (5-bromo-4-chloro-3-indolyl -D-glucuronide, 100 mM NaPO4 pH 7.5, 10 mM EDTA, 0.5 mM K3Fe(CN)6, 0.5 mM K4Fe(CN)6, 0.1% Titon X-100, 0.1% DMSO. After 12 h incubation at 37°C intact seedling were photographed. Alternatively, stained seedlings were vacuum infiltrated with Tissue-Tek OCT compound prior to freezing in liquid nitrogen. A bright cryostat microtome (model 5030) was used to cut 20 µm sections at -23°C.

Other modifications of the present invention will be apparent to those skilled in the art without departing from the scope of the invention.

REFERENCES

Ascenzi P, Aducci P, Torroni A, Amiconi G, Ballio A, Menegatti E, Guarneri M: The pH dependence of pre-steady-state and steady-state kinetics for the papain-catalyzed hydrolysis of n-alpha carbobenzoxyglycine p-nitrophenyl ester. Biochem Biophys Acta 912 (2): 203-210 (1987).

Becker C, Fischer J, Nong V H, Munitz K: PCR cloning and expression analysis of cDNAs encoding cysteine proteinases from germinating seeds of *Vicia sativa* L. Plant Mol Biol 26: 1207-1212 (1994).

Bevan, M: Binary *Agrobacterium* vectors for plant transformation. Nucleic Acids Res 12, 8711-8721 (1984).

Boylan M T, Sussex I M: Purification of an endopeptidase involved with storage-protein degradation in *Phaseolus vulgaris* L. cotyledons. Planta 170 (3): 343-352 (1987).

Bradford M M: A rapid and sensitive method for the quantification of microgram quantities of protein utilising the principle of protein-dye binding. Anal Biochem 72: 248-254 (1976).

Cammue B P A, De Bolle M F C, Terras F R G, Proost P, Van Damme J, Rees S B, Vanderleyden J, Broekaert W F: Isolation and characterisation of a novel class of plant antimicrobial peptides from *Mirabilis jalapa* L. Seeds. JBC 267 (4): 2228-2233 (1992).

Cervantes E, Rodriguez A, Nicolas G: Ethylene regulates the expression of a cysteine proteinase gene during germination of chickpea (*Cicer arietinum* L.). Plant Mol Biol 25 (2): 207-215 (1994).

Cercos M, Mikkonen A, Ho T-H D: Promoter analysis of a GA-induced cysteine endoprotease gene in barley aleurone cells. Plant Physiol (Rockville) 108 (2 SUPPL.): 79 (1995).

Cohen L W, Coghlan V M, Dihel L C: Cloning and sequencing of papain-encoding cDNA. Gene 48: 219-227 (1986).

Comai L, Harada J J: Transcriptionnal activities in dry seed nuclei indicate the timing of the transition from embryogeny to germination. Proc Natl Acad Sci USA 87: 2671-2674 (1990).

deBarros E G, Larkins B A: Cloning of a cDNA encoding a putative cysteine protease from germinating maize seeds. Plant Science 99 (2): 189-197 (1994).

Dietrich R A, Maslyar D J, Heupel R C, Harada J J: Spatial patterns of gene expression in *Brassica-napus* seedlings: identification of a cortex-specific gene and localization of messenger RNA encoding isocitrate lyase and a polypeptide homologous to proteinases. Plant Cell 1 (1): 73-80 (1989).

Edwards K, Johnstone C, Thompson C: A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res 19 (6): 1349 (1991).

Goetting-Minesky M P, Mullin B C: Differential gene expression in an actinorhizal symbiosis: Evidence for a nodule-specific cysteine proteinase. Proc Natl Acad Sci USA 91 (21): 9891-9895 (1994).

Goetting-Minesky P, Mullin B C: Evidence for a nodule-specific cysteine proteinase in an actinorhizal symbiosis. American Journal of Botany 81 (6 SUPPL.): 24 (1994).

Graham I A, Baker C J, Leaver C J: Analysis of the cucumber malate synthase gene promoter by transient expression and gel retardation assays. Plant Journal 6 (6): 893-902 (1994).

Grossberger: Minipreps of DNA from bacteriophage lambda . Nucleic Acids Res 15 (16): 6737 (1987).

Guerrero F D, Jones J T, Mullet J E: Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted sequence and expression of three inducible genes. Plant Mol Biol 15 (1): 11-26 (1990).

Hara-Nishimura I, Takeuchi Y, Nishimura M: Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of *Schistosoma mansoni*. Plant Cell 5 (11): 1651-1659 (1993).

Hara-Nishimura I, Shimada T, Hiraiwa N, Nishimura M: Vacuolar processing enzyme responsible for maturation of seed proteins. Journal of Plant Physiology 145 (5-6): 632-640 (1995).

Holsters M, de Waele D, Depicker A, Messen E, Van Montagu M, Schell J: Transfection and transformation of *A. tumefaciens*. Mol Gen Genet 163: 181-187 (1987).

Jefferson R A, Bevan M, Kavanagh T: The use of the *Escherichia coli* beta-glucuronidase gene as a gene fusion marker for studies of gene expression in higher plants - construction of promoter cloning vector plasmid pTAKI and detection of enzyme activity by fluorometric assay. Biochem Soc Trans 15: 17-18 (1987).

Jepson I, Bray J, Jenkins G, Schuch W, Edwards K: A rapid procedure for the construction of PCR cDNA libraries from small amounts of plant tissue. Plant Mol Biol Reporter 9 (2): 131-138 (1991).

Jiang B, Siregar U, Willeford K O, Luthe D S, Williams W P: Association of a 33-kilodalton cysteine proteinase found in corn callus with the inhibition of fall armyworm larval growth. Plant Physiol 108 (4): 1631-1640 (1995).

Jones M L, Larsen P B, Woodson W R: Ethylene-regulated expression of a carnation cysteine proteinase during flower petal senescence. Plant Mol Biol 28 (3): 505-512 (1995).

Karrer K M, Peiffer S L, DiTomas M E: Two distinct gene subfamilies within the family of cysteine protease genes. Proc Natl Acad Sci USA 90: 3063-3067 (1993).

Kianian S F, Quiros C F: Genetic analysis of major multigene families in *Brassica oleracea* and related species. Genome 35 (3): 516-527 (1992).

Koehler S M, Ho T-H D: Hormonal regulation processing and secretion of cysteine proteinases in barley aleurone layers. Plant Cell 2 (8): 769-784 (1990).

Korodi I; Asboth B; Polgar L: Disulfide bond formation between the active-site thiol and one of the several free thiol groups of chymopapain. Biochemistry 25 (22): 6895-6900 (1986).

Lidgett A, Moran M, Wong K A L Furze J, Rhodes M J C and Hamill J D: Isolation and expression pattern of a cDNA encoding a cathepsin B-like protease from *Nicotiana rustica*. Plant Mol Biol 29 (2): 379-384 (1995).

Lin E; Burns D J W; Gardner R C: Fruit developmental regulation of the kiwifruit actinidin promoter is conserved in transgenic petunia plants. Plant Mol Biol 23 (3): 489-499 (1993).

Linthorst H J M, Van Der Does C, Van Kan J A L, Bol J F: Nucleotide sequence of a cdna clone encoding tomato *Lycopersicon esculentum* cysteine proteinase. Plant Physiol 101 (2): 705-706 (1993).

Linthorst H J M, Van Der Does C, Brederode F T, Bol J F: Circadian expression and induction by wounding of tobacco genes for cysteine proteinase. Plant Mol Biol 21 (4): 685-694 (1993).

Mallinson D J, Lockwood B C, Coombs G H, North M J: Identification and molecular cloning of four cysteine proteinase genes from the pathogenic protozoon *Trichomonas vaginalis*. Microbiology 140 (10): 2725-2735 (1994).

Mikkelsen T R, Andersen B and Jorgensen R B: The risk of crop transgene spread. Nature 380: 31 (7 March 1996).

Marttila S, Porali I, Ho T-H D, Mikkonen A: Expression of the 30 kd cysteine endoprotease b in germinating barley seeds. Cell biol int 17 (2): 205-212 (1993).

Marttila S, Jones B L, Mikkonen A: Differential localization of two acid proteinases in germinating barley (*Hordeum vulgare*) seed. Physiologia Plantarum 93 (2): 317-327 (1995).

Mckee R A, Adams S, Matthews J A, Smith C J, Smith H: Molecular cloning of two cysteine proteinases from paw-paw (carica-papaya). Biochem J 237 (1): 105-1101 (1986).

Minami A and Fukuda H: Transient and specific expression of a cysteine endopeptidase associated with autolysis during differentiation of *Zinnia* mesophyll cells into tracheary elements. Plant Cell Physiol 36 (8): 1599-1606 (1995).

Mino M, Inoue M: Hybrid vigor in relation to lipid and protein metabolism in germinating maize kernels. Jpn J Breed 38 (4): 428-436 (1988).

Nong V H, Becker C, Muentz K: cDNA cloning for a putative cysteine proteinase from developing seeds of soybean. Biochim et Biophys Acta 1261 (3): 435-438 (1995).

Okayama H, Kawaichi M, Brownstein M, Lee F, Yokota T, Arai K: High-efficiency cloning of full-length cDNA: construction and screening of cDNA expression libraries from mammalian cells. Meth. in Enzymology 154: 3-27 (1987).

Pautot V, Brzezinski R, Tepfer M: Expression of a mouse metallothionein gene in transgenic plant tissues. Gene 77:133-140 (1989).

Pladys D, Vance C P: Proteolysis during development and senescence of effective and plant gene-controlled ineffective alfalfa nodules. Plant Physiol 103 (2): 379-384 (1993).

Praekelt U M, Mckee R A, Smith H: Molecular analysis of actinidin the cysteine proteinase of *Actinidia chinensis*. Plant Mol Biol 10 (3): 193-202 (1988).

Revell d f, Cummings N J, Baker K C, Collins M E, Taylor M A J, Sumner I G, Pickersgill R W, Connerton I F, Goodenough P W: Nucleotide sequences and expression in *Escherichia coli* of cDNA encoding papaya proteinase omega from *Carica papaya*. Gene 127 (2): 221-225 (1993).

Sambrook J, Fritsch E F, Maniatis T: Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).

Sanger F, Milkin S, Coulson A R: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5467 (1977).

Schaffer M A, Fischer R L: Analysis of mRNAs that accumulate in response to low temperatures identifies a thiol protease gene in tomato. Plant Physiol 87: 431-436 (1988).

Shimada T,Hiraiwa N, Nishimura M, Hara-Nishimura I: Vacuolar processing enzyme of soybean that converts proproteins to the corresponding mature forms. Plant Cell Physiol 35 (4): 713-718 (1994).

Shintani A, Yamauchi D, Minamikawa T: Nucleotide sequence of cDNA for a putative cysteine protease from rice seeds. Plant Physiol 107 (3): 1025 (1995).

Shutov A D, Vaintraub I A: Degradation of storage proteins in germinating seeds. Phytochemistry 26 (6): 1557-1566 (1987).

Smart C M, Hosken S E, Thomas H, Greaves J A, Blair B G, Schuch W: The timing of maize leaf senescence and characterisation of senescence-related cDNAs. Physiol Plant 93 (4): 673-682 (1995).

Takeda O, Miura Y, Mitta M, Matsushita H, Kato I, Abe Y, Yokosawa H Ishii S-I: Isolation and analysis of cDNA encoding a precursor of *Canavalia ensiformis* asparaginyl endopeptidase (Legumain). Journal of Biochemistry 16 (3): 541-546 (1994).

Taylor M A J, Baker K C, Briggs G S, Connerton I F, Cummings N J, Pratt K A, Revell D F, Freedman R B, Goodenough P W: Recombinant pro-regions from papain and papaya proteinase IV are selective high affinity inhibitors of the mature papaya enzymes. Protein Engineering 8 (1): 59-62 (1995).

Terras R G, Torrekens S, Van Leuven F, Osborn R W, Vanderleyden J, Cammue B P A, Broekaert W F: A new family of basic cysteine-rich plant antifungal proteins from *Brassicaceae* species. FEBS Lett 316 (3): 233-240 (1993).

Thomas M P, Topham C M, Kowlessur D, Mellor G W, Thomas E W, Whitford D, Brocklehurst K: Structure of chymopapain M the late-eluted chymopapain deduced by comparative modelling techniques and active-centre characteristics determined by pH-dependent kinetics of catalysis and reactions with time-dependent inhibitors: The Cys-25-His-

159 ion-pair is insufficient for catalytic competence in both chymopapain M and papain. Biochemical Journal 300 (3): 805-820 (1994).

Valpuesta V, Lange N E, Guerrero C, Reid M S: Up-regulation of a cysteine protease accompanies the ethylene-insensitive senescence of daylily (Hemerocallis) flowers. Plant Mol Biol 28 (3): 575-582 (1995).

Vernet T, Berti P J, De Montigny C, Musil R, Tessier D C, Menard R, Magny M-C, Storer A C, Thomas D Y: Processing of the papain precursor: the ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. J Biol Chem 270 (18): 10838-10846 (1995).

Watanabe H, Abe K, Emori Y, Hosoyama H, Arai S: Molecular cloning and gibberellin-induced expression of multiple cysteine proteinases of rice seeds (oryzains). J Biol Chem 266 (25): 16897-16902 (1991).

Wiederanders B, Broemme D, Kirschke H, Von Figura K, Schmidt B, Peters C: Phylogenetic conservation of cysteine proteinases cloning and expression of a cDNA coding for human cathepsins. J Biol Chem 267 (19): 13708-13713 (1992).

Yamauchi D, Akasofu H, Minamikawa T: Cysteine endopeptidase from *Vigna mungo*: gene structure and expression. Plant Cell Physiol. 33 (6): 789-797 (1992).

Claims

1. A promoter comprising the DNA sequence of an oil seed rape cysteine protease gene promoter of class 1, 2 or 6.

2. A promoter comprising at least part of the DNA sequence as shown in Figure No.19, or at least part of a sequence that has substantial homology therewith, or a variant thereof.

3. A promoter comprising at least part of the DNA sequence as shown in Figure No. 20 or at least part of a sequence that has substantial homology therewith, or a variant thereof.

4. A promoter comprising at least part of the DNA sequence as shown in Figure 21 or at least part of a sequence that has substantial homology therewith, or a variant thereof.

5. A promoter having the characteristic motif or features of a promoter of any one of the preceding claims.

6. A recombinant DNA construct comprising the promoter of any preceding claim operably linked to a gene of interest.

7. A recombinant DNA construct functional in a plant comprising a disrupter gene encoding a product capable of disrupting cell function, and a promoter according to any preceding claim for the disrupter gene, the disrupter gene being functionally linked to and controlled by an externally regulatable gene control region which includes a promoter which is inducible by the external application of a chemical inducer.

8. A recombinant DNA construct according to claim 7 wherein the inducible promoter is functionally linked to and controls a repressor protein gene and in which the disrupter gene promoter includes an operator sequence which is recognised by the repressor

protein, so that in the presence of the inducer the repressor protein is produced which interacts with the operator sequence thereby disabling the second promoter and inhibiting expression of the disrupter gene.

9. A recombinant DNA construct according to claim 7 or claim 8 wherein the disrupter gene is a nucleotide sequence, which is in sense orientation to an endogenous plant gene which is essential to plant development or a gene conferring a desired characteristic on the plant, or comprises a partial sense sequence of the endogenous plant gene.

10. A recombinant DNA construct according to claim 7 or claim 8 wherein the disrupter gene is a nucleotide sequence, which is in antisense orientation to an endogenous plant gene which is essential to plant development or a gene conferring a desired characteristic on the plant.

11. A recombinant DNA construct according to claim 9 or claim 10 wherein the endogenous plant gene is essential to seed germination or early seedling development.

12. A recombinant DNA construct according to any one of claims 7 to 11 wherein the externally regulatable gene control region is a chemically inducible gene promoter sequence from the glutathione S-transferase system, the *Alc* system or the ecdysone system.

13. A recombinant DNA construct according to any one of claims 8 to 13 wherein the repressor protein gene encodes a bacterial repressor.

14. A recombinant DNA construct according to claim 13 wherein the repressor protein gene encodes the *lac* repressor or a repressor used by 434, P22 or lambda-bacteriophages.

15. A recombinant DNA construct according to any one of claims 7 to 14 wherein the disrupter gene or disrupter promoter contains a "pseudo-operator".

16. A recombinant DNA construct according to any one of claims 7 to 15 wherein the disrupter gene is a cytotoxic gene.

17. A recombinant DNA construct according to any one of claims 7 to 15 wherein the disrupter gene encodes a recombinase or transposase adapted to excise a nucleotide sequence flanked by recombinase recognition sequences.

18. DNA comprising at least part of the sequence shown in Figure 12 or at least part of a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.

19. DNA comprising at least part of the sequence shown in Figure 13 or at least part of a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.

20. DNA comprising at least part of the sequence shown in Figure 14 or at least part of a sequence that has substantial homology therewith or a variant thereof and which codes for a cysteine protease.

21. DNA comprising at least part of the sequence shown in Figure 15 or at least part of a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.

22. DNA comprising at least part of the sequence shown in Figure 16 or at least part of a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.

23. DNA comprising at least part of the sequence shown in Figure 17 or at least part of a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.

24. A recombinant DNA construct functional in a plant comprising the DNA of any one of claims 18 to 23 operably linked to a promoter.

25. A recombinant DNA construct according to any one of claims 6 to 17 and 24 wherein the construct is capable of being expressed in the tissue or tissues of a germinating seedling or a developing grain or a plant when the construct is integrated , preferably stably integrated, within the grain's or seedling's or plant's genomic DNA.

26. An expression system for the tissue or tissues of a plant material, the expression system comprising
a gene of interest fused to a gene promoter as defined in any one of claims 1 to 5 wherein the expression system is capable of being expressed in the tissue or tissues of the plant material.

27. An expression system according to claim 26 wherein the expression system is for at least the tissue of a germinating seedling or developing grain or plant (eg in the root, cotyledons, leaves and stem).

28. An expression system according to claim 26 or claim 27 wherein the expression system is integrated, preferably stably integrated, within a germinating seedling's genomic DNA or a developing grain's genomic DNA or a plant's genomic DNA.

29. An expression system comprising a construct according to any one of claims 6 to 17 and 24.

30 A recombinant plant genome comprising a promoter of any one of claims 1 to 5, DNA of any one of claims 18 to 23, a recombinant DNA construct of any one of claims 6 to 17 or 24 or an expression system according to any one of claims 25 to 29.

5 31. A plant, plant seed or plant cell having a recombinant plant genome of claim 30.

32. Protected plant germplasm comprising a plant comprising a recombinant DNA construct of any one of claims 6 to 17 or 24.

10 33. A plant or seed which is incapable of growing to maturity comprising a recombinant DNA construct of any one of claims 6 to 17 or 24.

34. The use of a gene promoter as defined in any one of claims 1 to 5 to induce expression of a gene of interest when fused to the gene promoter in the tissue or tissues of a plant
15 material.

35. The use according to claim 34 wherein the gene promoter is used to induce expression of a gene of interest when fused to the gene promoter in at least the tissue or tissues of a germinating seedling or a developing grain or a plant (eg in the root, cotyledons,
20 leaves and stem).

36. A promoter, a construct or an expression system substantially as hereinbefore described with reference to any one of Figures 12 to 17 and 19 to 21.

25

Fig.1.

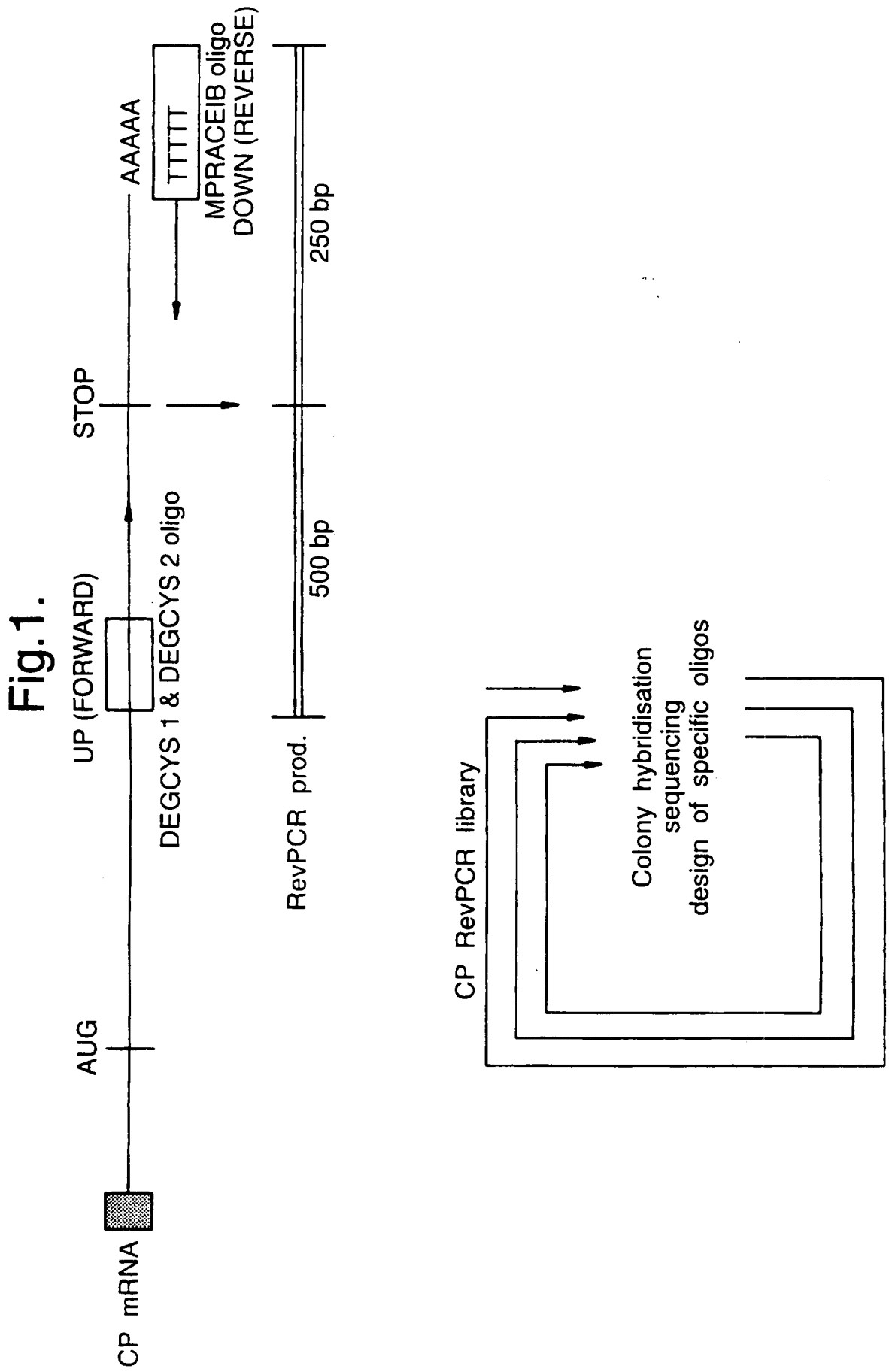

CP mRNA
AUG
UP (FORWARD)
DEGCYS 1 & DEGCYS 2 oligo
STOP
AAAAA
TTTTT
MPRACEIB oligo
DOWN (REVERSE)
RevPCR prod.
500 bp
250 bp
CP RevPCR library
Colony hybridisation
sequencing
design of specific oligos

Fig.2.

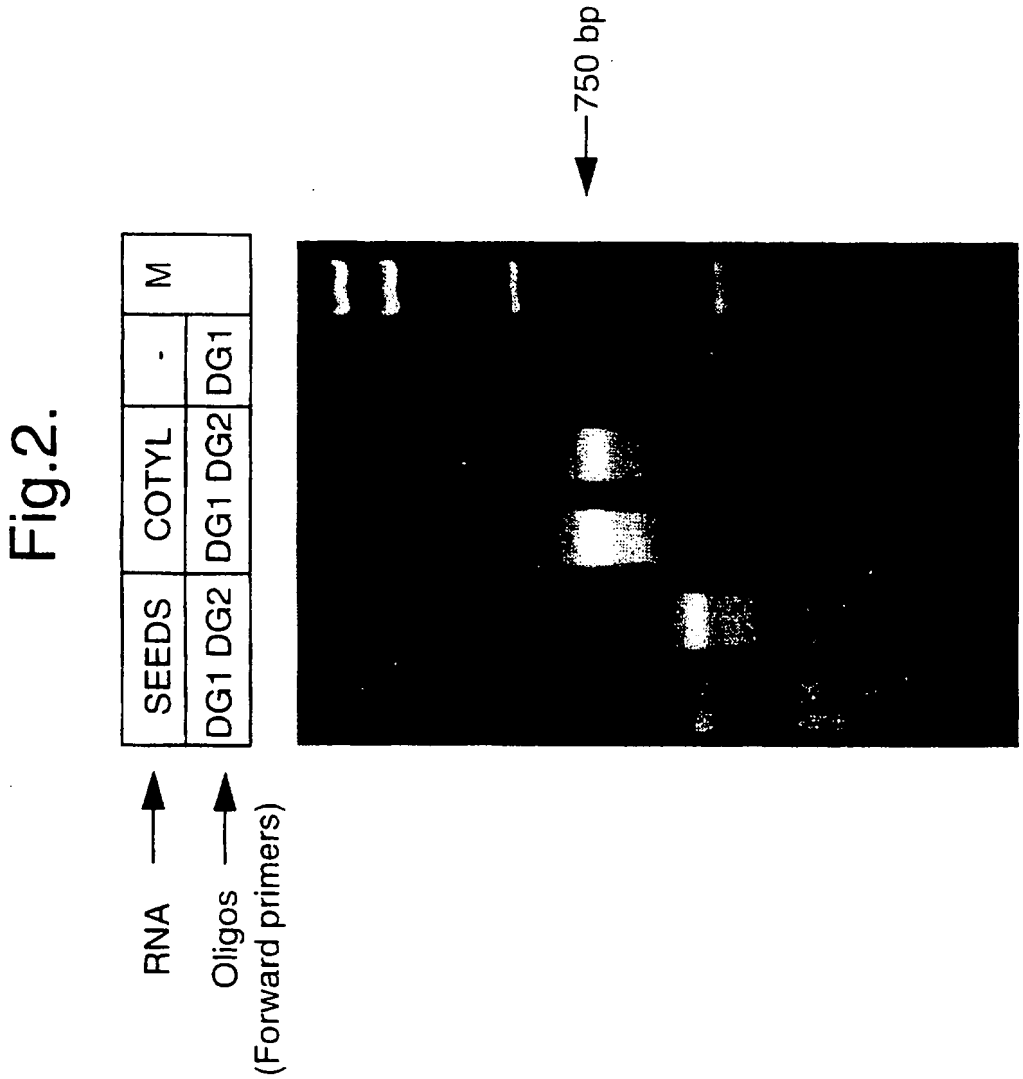

750 bp
SEEDS
COTYL
-
M
DG1 DG2
DG1 DG2
DG1
RNA
Oligos
(Forward primers)

Fig.3A. 1/8

Solution Parameters:

```
Nucleic Alphabet    = Identity
Output line length  = 80
Compress            = Off
Histogram           = Off
Randomization       = Off

AMINO-Res-length    = 2
DELetion-weight     = 10.00
LEngth-factor       = 0
Matching-weight     = 10.00
NUCLEIC-Res-length  = 4
SPread-factor       = 90
```

Clustered order of selected sequences:

			with 8.403	
165/288	1.	OSR8.401COD	57% humd.	(1-317)
	4.	OSR8.406COD		(1-361)
	5.	OSR8.403COD		(1-468)
	15.	OSR8.404COD		(1-501)
	16.	OSR8.402COD		(1-501)
127/474	13.	OSR8.389COD	90% humd.	(1-504)
	12.	OSR8.387COD		(1-504)

Fig.3A. 2/8

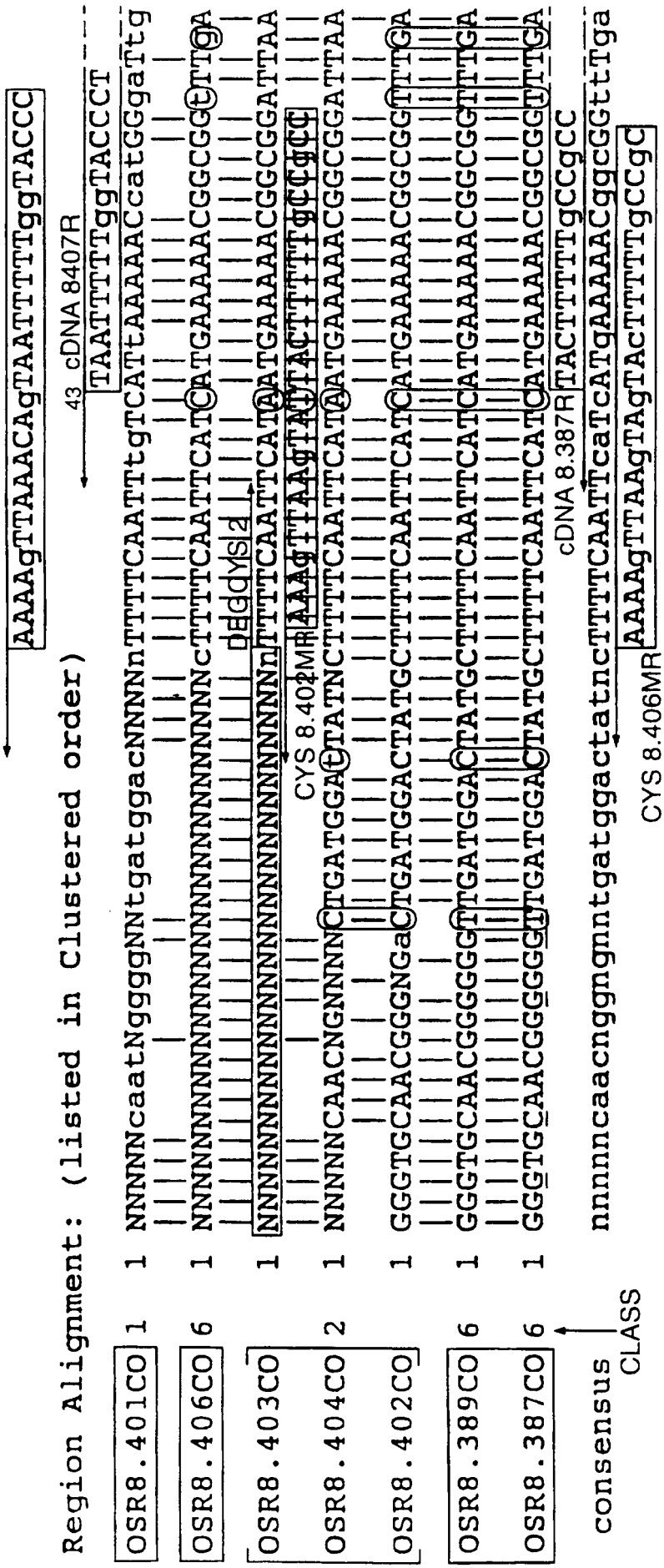

Region Alignment: (listed in Clustered order)
CYS 8.407MR
AAAAgTTAAACAgTAATTTTTggTACCC
43 cDNA 8407R
TAATTTTTggTACCCT
OSR8.401CO 1
OSR8.406CO 6
OSR8.403CO
OSR8.404CO 2
OSR8.402CO
OSR8.389CO 6
OSR8.387CO 6
consensus
CLASS
CYS 8.402MR
CYS 8.406MR
cDNA 8.387R

Fig.3A. 3/8

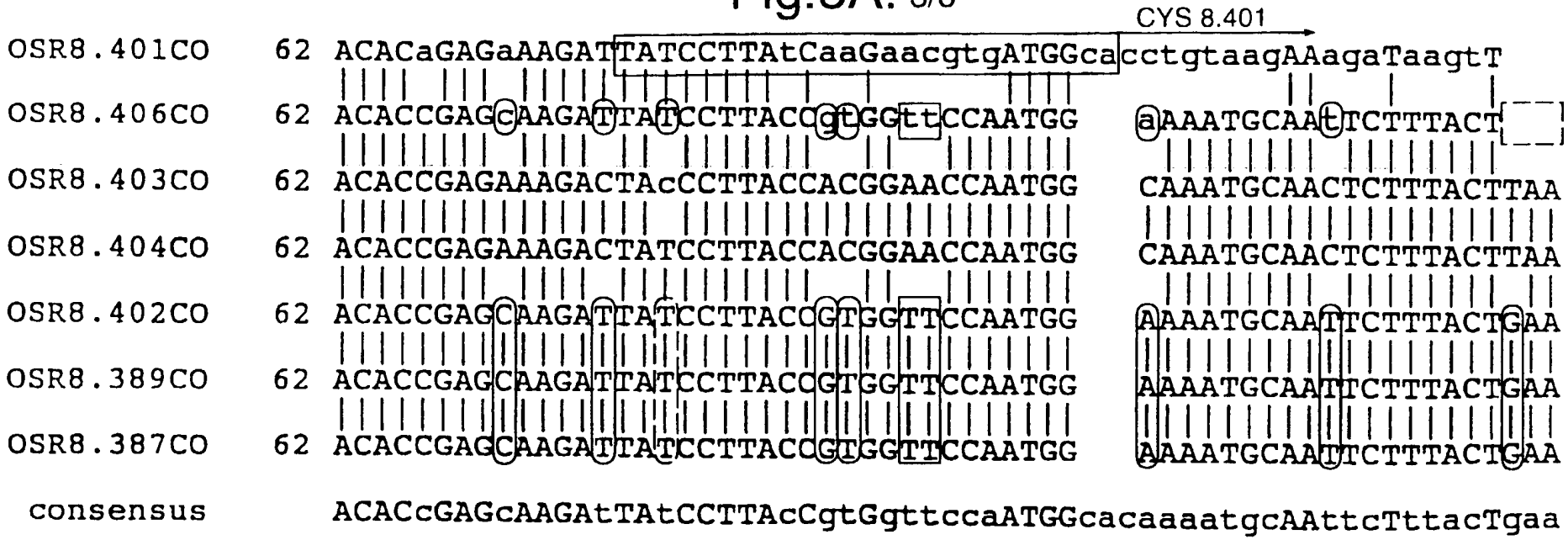


Fig.3A. 4/8

6/89

Fig.3A. 5/8

Annotations: 203, CYS, 8.406, 225, 232

OSR8.401CO	181	TTactagaAGCtGTagncgctCAGCCaGTtAGTGTTGgtATctgtGggaGcGagAGAGcgT
OSR8.406CO	178	TTGAAGAGAGCAGTTTCATACCAGCCCGTGAGTGTTGCCATTGAaGCTGGTGGAAGAGtTT
OSR8.403CO	181	TTGAAGAGAGCAGTTTCATACCAGCCTGTGAGTGTTGCTATTGATGCTGGTGGAAGAGCTT
OSR8 404CO	181	TTGAAGAGAGCAGTTTCgTACCAGCCTGTGAGTGTTGCTATTGATGCTGGTGGAAGAGCTT
OSR8.402CO	181	TTGAAGAGAGCAGTTTCATACCAGCCTGTGAGTGTTnCTATTGATnCTGGTGGAAnnnCTT
OSR8.389CO	181	TTGAAGAGAGCAGTTTCATACCAGCCcGTGAGTGTTgCcATTGAagCTGGTGGAAgagTTT
OSR8.387CO	181	nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnTTT
consensus		ttgaagagagcagtttcataccagcctgtgagtgttgctattgatgctggtggaagagctT

Fig.3A. 6/8

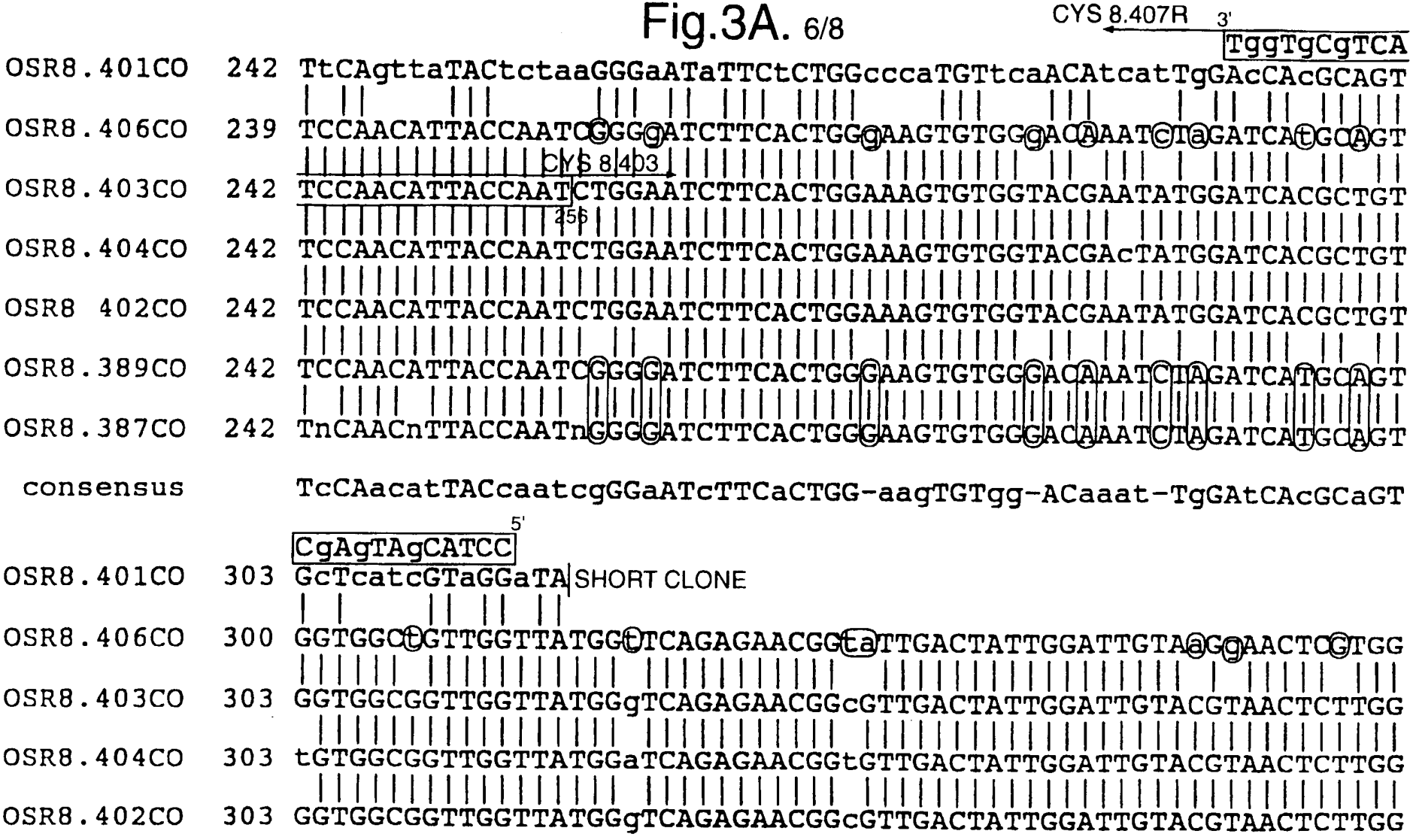


Fig.3A. 7/8

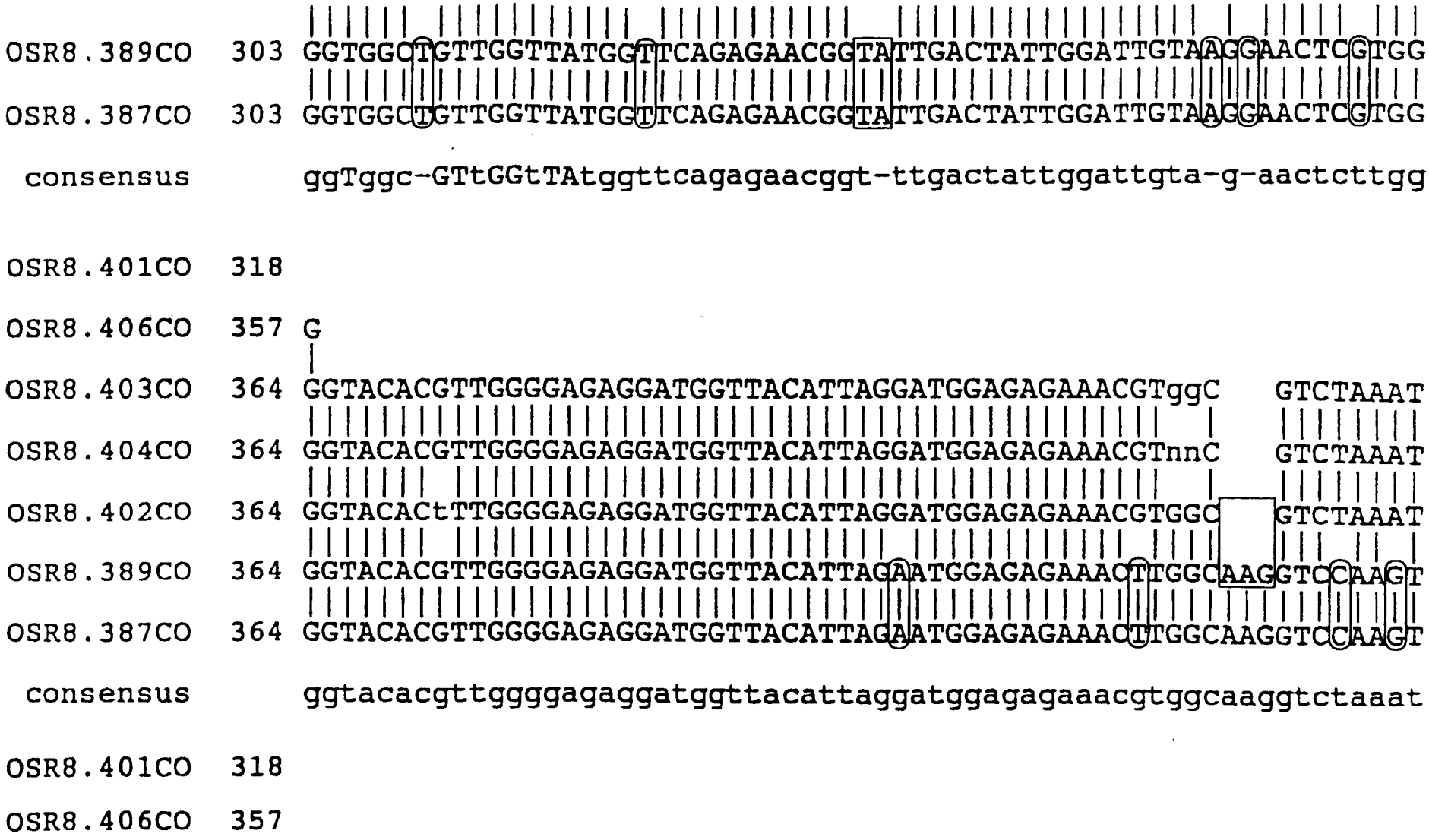
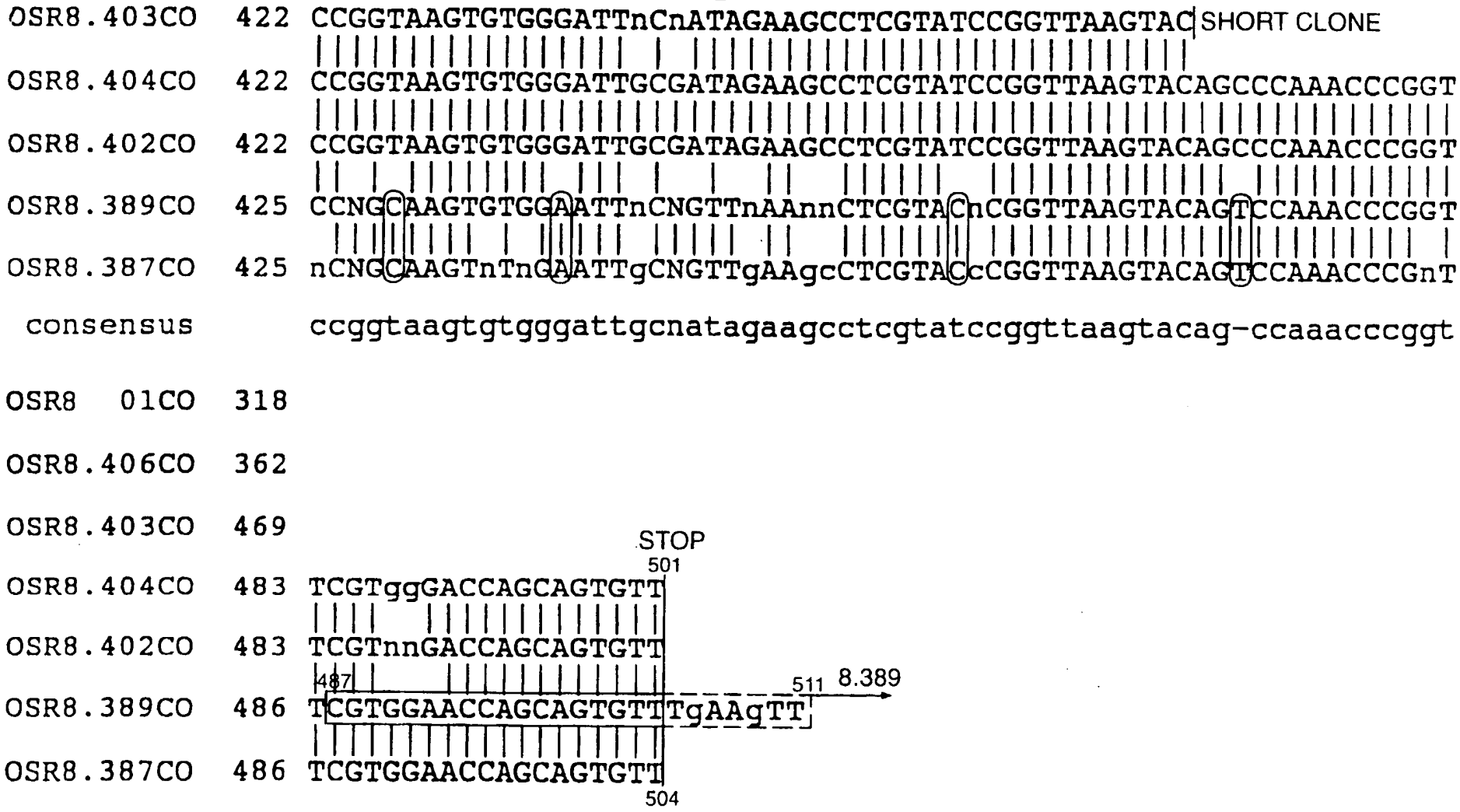


Fig.3A. 8/8



Alignment score = 19760.00

Fig.3B. 1/3

```
DELetion-weight     = 1.00
LEngth-factor       = 0
Matching-weight     = 5.00
NUCLEIC-Res-length  = 4
SPread-factor       = 90

Clustered order of selected sequences:

    14. OSR8.389NCOD  75% humd. (1-240)
    17. OSR8.387NCOD  with 8.402 (1-246)
     7. OSR8.402NCOD      8.404 (1-242)
     6. OSR8.404NCOD      8.403 (1-242)

Region Alignment: (listed in Clustered order)
```

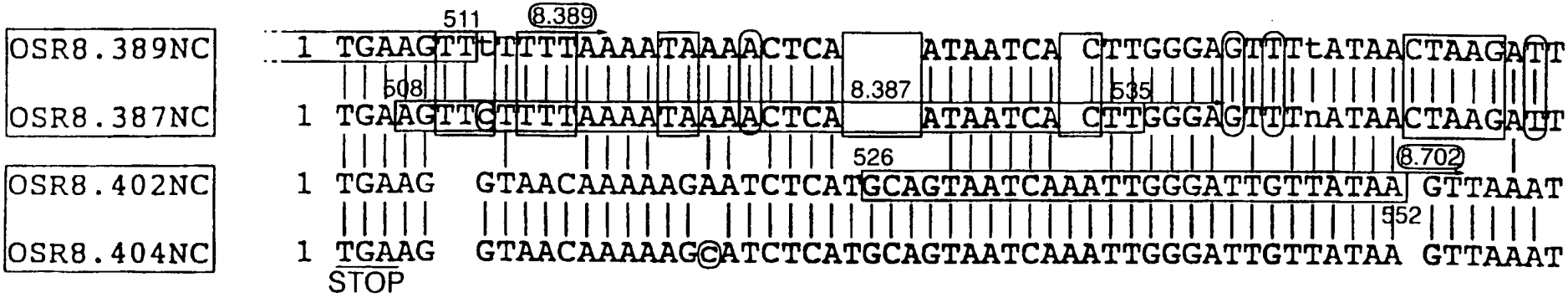


Fig.3B. 2/3

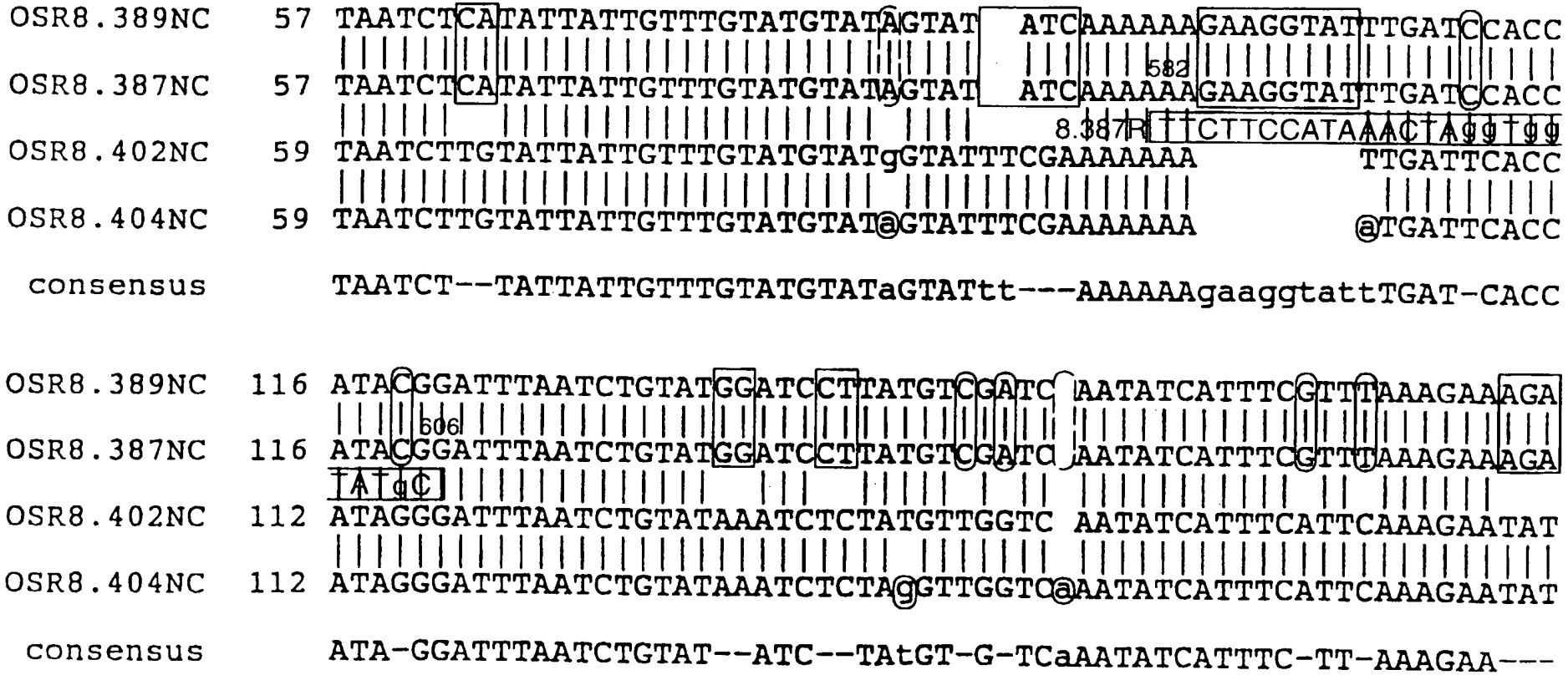

OSR8.389NC 57 TAATCTCATATTATTGTTTGTATGTATAGTAT ATCAAAAAAGAAGGTATTTGATCCACC
582
OSR8.387NC 57 TAATCTCATATTATTGTTTGTATGTATAGTAT ATCAAAAAAGAAGGTATTTGATCCACC
8.387R TTCTTCCATAAACTAGGTGG
OSR8.402NC 59 TAATCTTGTATTATTGTTTGTATGTATgGTATTTCGAAAAAAA TTGATTCACC
OSR8.404NC 59 TAATCTTGTATTATTGTTTGTATGTAT@GTATTTCGAAAAAAA @TGATTCACC
consensus TAATCT--TATTATTGTTTGTATGTATaGTATTt---AAAAAAgaaggtattTGAT-CACC
OSR8.389NC 116 ATACGGATTTAATCTGTATGGATCCTTATGTCGATC AATATCATTTCGTTTAAAGAAAGA
606
OSR8.387NC 116 ATACGGATTTAATCTGTATGGATCCTTATGTCGATC AATATCATTTCGTTTAAAGAAAGA
TATgC
OSR8.402NC 112 ATAGGGATTTAATCTGTATAAATCTCTATGTTGGTC AATATCATTTCATTCAAAGAATAT
OSR8.404NC 112 ATAGGGATTTAATCTGTATAAATCTCTAGGTTGGTC@AATATCATTTCATTCAAAGAATAT
consensus ATA-GGATTTAATCTGTAT--ATC--TAtGT-G-TCaAATATCATTTC-TT-AAAGAA---

Fig.3B. 3/3

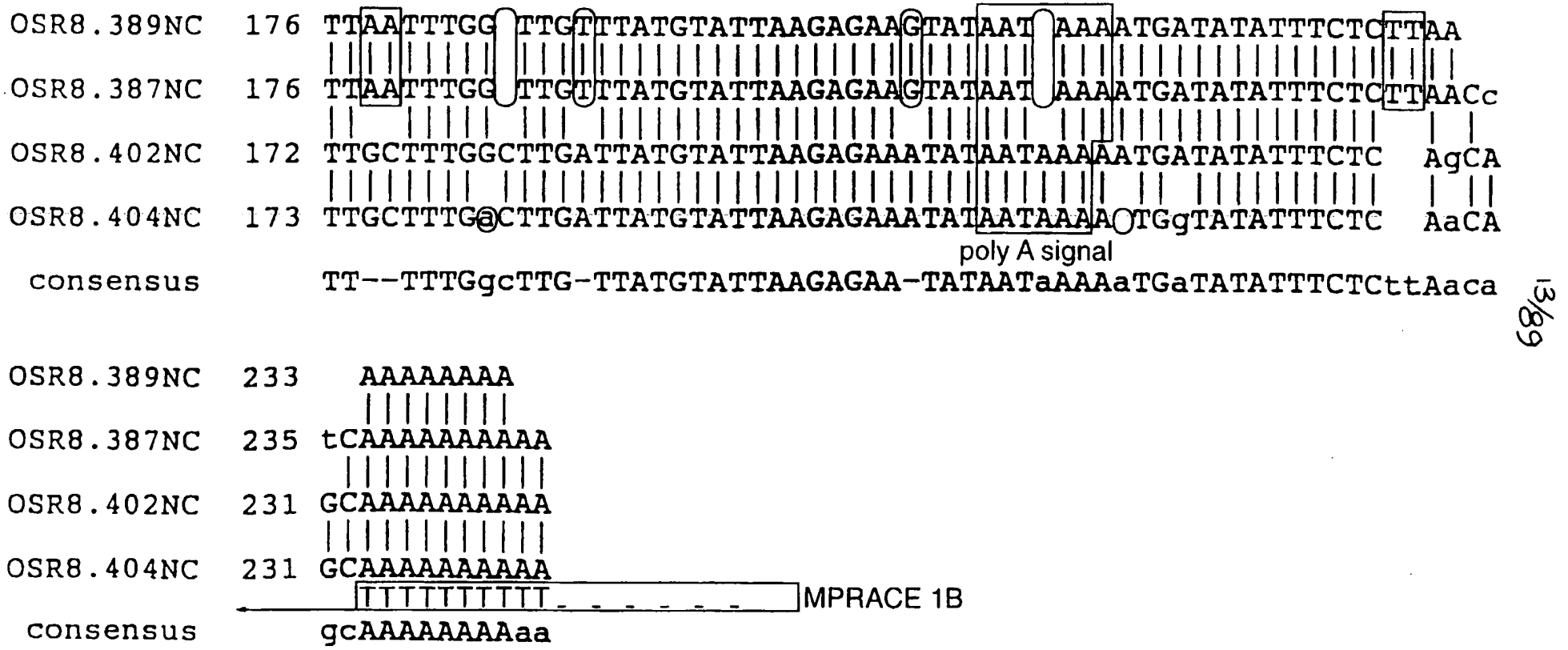


Fig.4. 1/4

OSR8.401COD from 1 to 317: (CLASS 1)

```
       10          20          30          40          50          60          70
nnnnncaatn ggggnntgat ggacnnnnnt tttcaatttg tcattaaaaa ccatgggatt gacacagaga

       80          90         100         110         120         130         140
aagattatcc ttatcaagaa cgtgatggca cctgtaagaa agataagttg aatagaaagg ttgtgacaat

      150         160         170         180         190         200         210
tgatagctac gctggtgtaa aatcaaatga cgagaaagcg ttactagaag ctgtagncgc tcagccagtt

      220         230         240         250         260         270         280
agtgttggta tctgtgggag cgagagagcg tttcagttat actctaaggg aatattctct ggcccatgtt

      290         300         310
caacatcatt ggaccacgca gtgctcatcg taggata
```

GEL: pri 40

Range to print (<CR>=ALL):
OSR8-402 from 1 to 743: (CLASS 2)

```
       10          20          30          40          50          60          70
gggtgcaacg ggngactgat ggactatgct tttcaattca tcatgaaaaa cggcggtttg aacaccgagc
```

14/89

Fig.4. 2/4

```
        80          90         100         110         120         130         140
aagattatcc ttaccgtggt tccaatggaa aatgcaattc tttactgaag aattcaagag ttgtaactat

       150         160         170         180         190         200         210
tgatggatac gaagatgttc ctagtaaaga tgaaaccncn ttgaagagag cagtttcata ccagcctgtg

       220         230         240         250         260         270         280
agtgttncta ttgatnctgg tggaannnct ttccaacatt accaatctgg aatcttcact ggaaagtgtg

       290         300         310         320         330         340         350
gtacgaatat ggatcacgct gtggtggcgg ttggttatgg gtcagagaac ggcgttgact attggattgt

       360         370         380         390         400         410         420
acgtaactct tggggtacac tttggggaga ggatggttac attaggatgg agagaaacgt ggcgtctaaa

       430         440         450         460         470         480         490
tccggtaagt gtgggattgc gatagaagcc tcgtatccgg ttaagtacag cccaaacccg gttcgtnnga

       500         510         520         530         540         550         560
ccagcagtgt ttgaaggtaa caaaaagaat ctcatgcagt aatcaaattg ggattgttat aagttaaatt

       570         580         590         600         610         620         630
aatcttgtat tattgtttgt atgtatggta tttcgaaaaa aattgattca ccatagggat ttaatctgta

       640         650         660         670         680         690         700
taaatctcta tgttggtcaa tatcatttca ttcaaagaat atttgctttg gcttgattat gtattaagag

       710         720         730         740
aaatataata aaaatgatat atttctcagc agcaaaaaaa aaa
```

Fig.4. 3/4

CYS8-389 from 1 to 744: (CLASS 6)

```
        10        20        30        40        50        60        70
gggtgcaacg ggggggttgat ggactatgct tttcaattca tcatgaaaaa cggcgggtttg aacaccgagc

        80        90       100       110       120       130       140
aagattatcc ttaccgtggt tccaatggaa aatgcaattc tttactgaag aattcaagag ttgtaactat

       150       160       170       180       190       200       210
tgatggttac gaagatgttc ctactgaaga tgaaaacggcg ttgaagagag cagtttcata ccagcccgtg

       220       230       240       250       260       270       280
agtgttgcca ttgaagctgg tggaagagtt ttccaacatt accaatcggg gatcttcact gggaagtgtg

       290       300       310       320       330       340       350
ggacaaatct agatcatgca gtggtggctg ttggttatgg ttcagagaac ggtattgact attggattgt
```

Fig.4. 4/4

```
                360       370       380       390       400       410       420
aaggaactcg tggggtacac gttgggggaga ggatggttac attagaatgg agagaaactt ggcaaggtcc

                430       440       450       460       470       480       490
aagtccngca agtgtggaat tncngttnaa nnctcgtacn cggttaagta cagtccaaac ccggttcgtg

                500       510       520       530       540       550       560
gaaccagcag tgtttgaagt ttttttaaaa taaaactcaa taatcacttg ggagttttat aactaagatt

                570       580       590       600       610       620       630
taatctcata ttattgtttg tatgtatagt atatcaaaaa agaaggtatt tgatccacca tacggattta

                640       650       660       670       680       690       700
atctgtatgg atccttatgt cgatcaatat catttcgttt aaagaaagat taatttggtt gtttatgtat

                710       720       730       740
taagagaagt ataataaaat gatatatttc tcttaaaaaa aaaa
```

Fig.5.

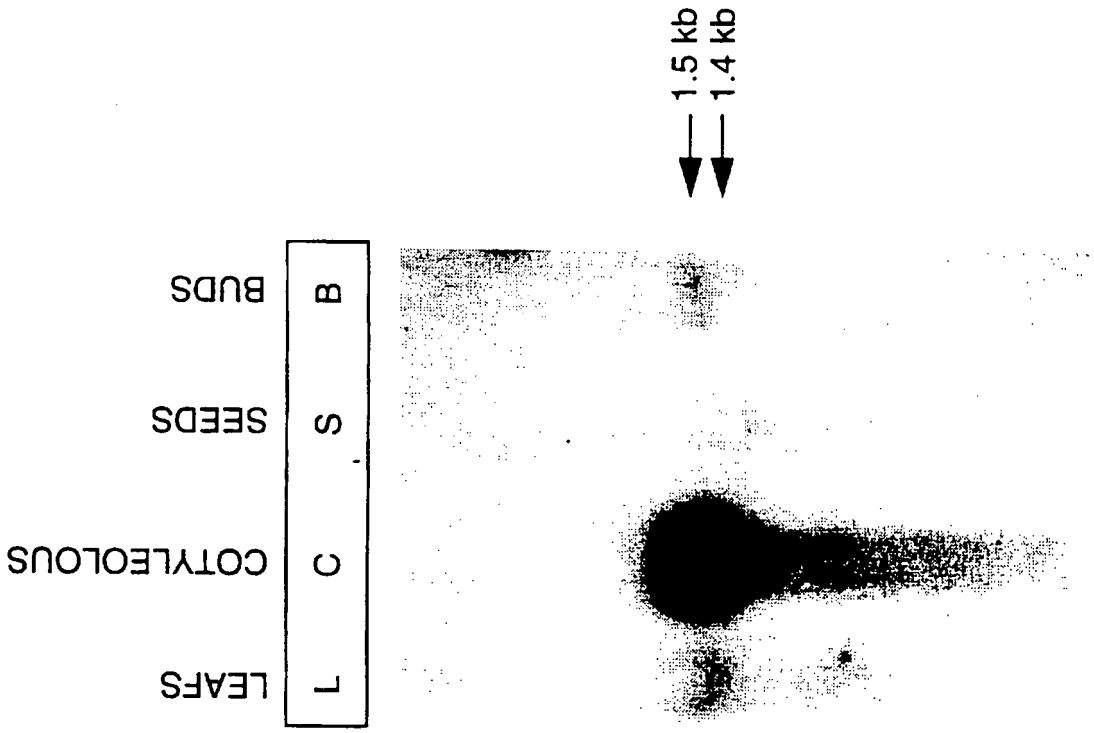

1.5 kb
1.4 kb
BUDS
SEEDS
COTYLEDOLOUS
LEAFS
B
S
C
L
CLASS 2
Solution primed
whole RT-PCR fragment
used as a probe

Fig.6.

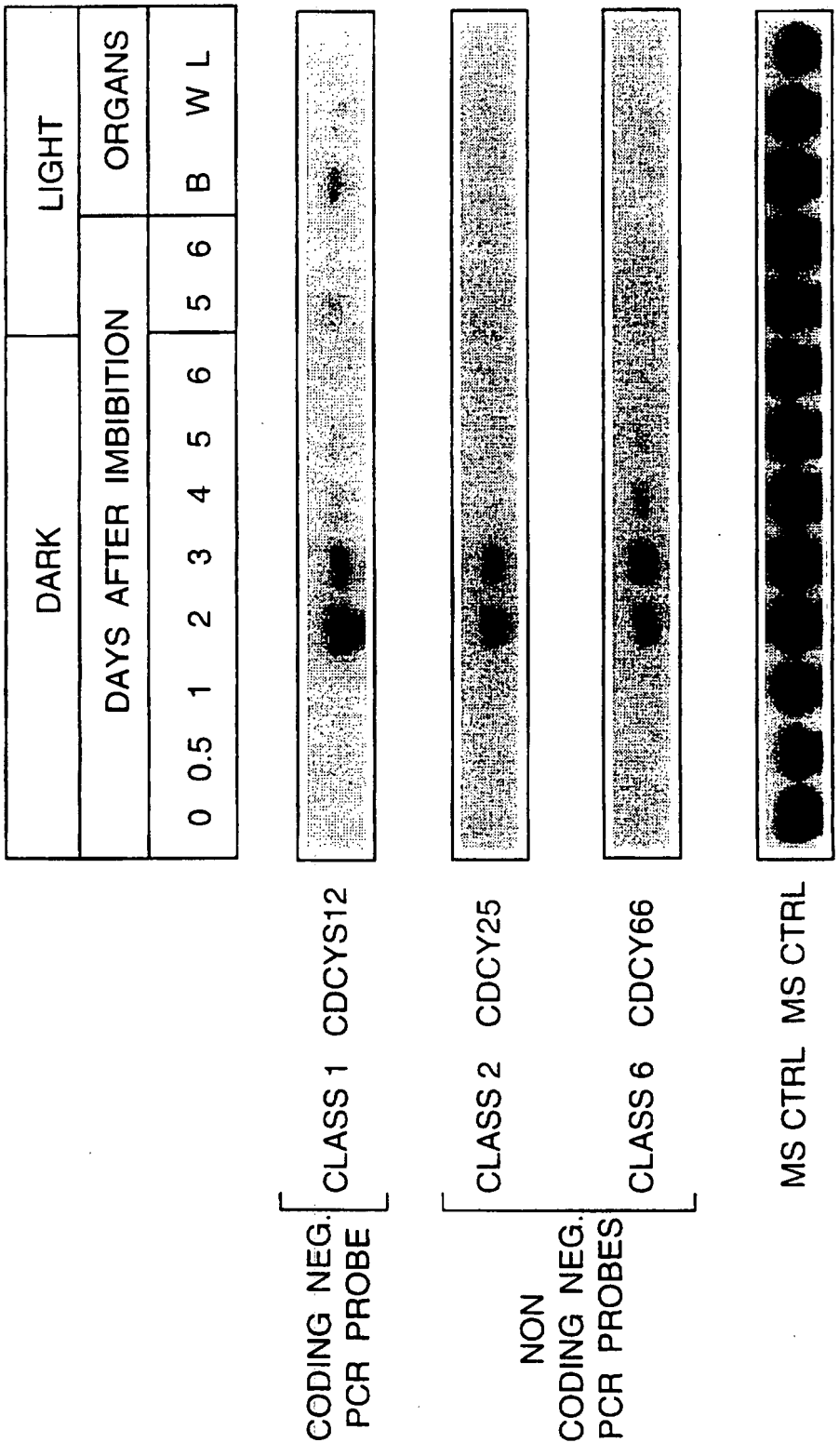

DARK
LIGHT
DAYS AFTER IMBIBITION
0 0.5 1 2 3 4 5 6
5 6
ORGANS
B W L
CODING NEG. PCR PROBE
CLASS 1 CDCYS12
NON CODING NEG. PCR PROBES
CLASS 2 CDCY25
CLASS 6 CDCY66
MS CTRL MS CTRL

Fig.7. 1/4

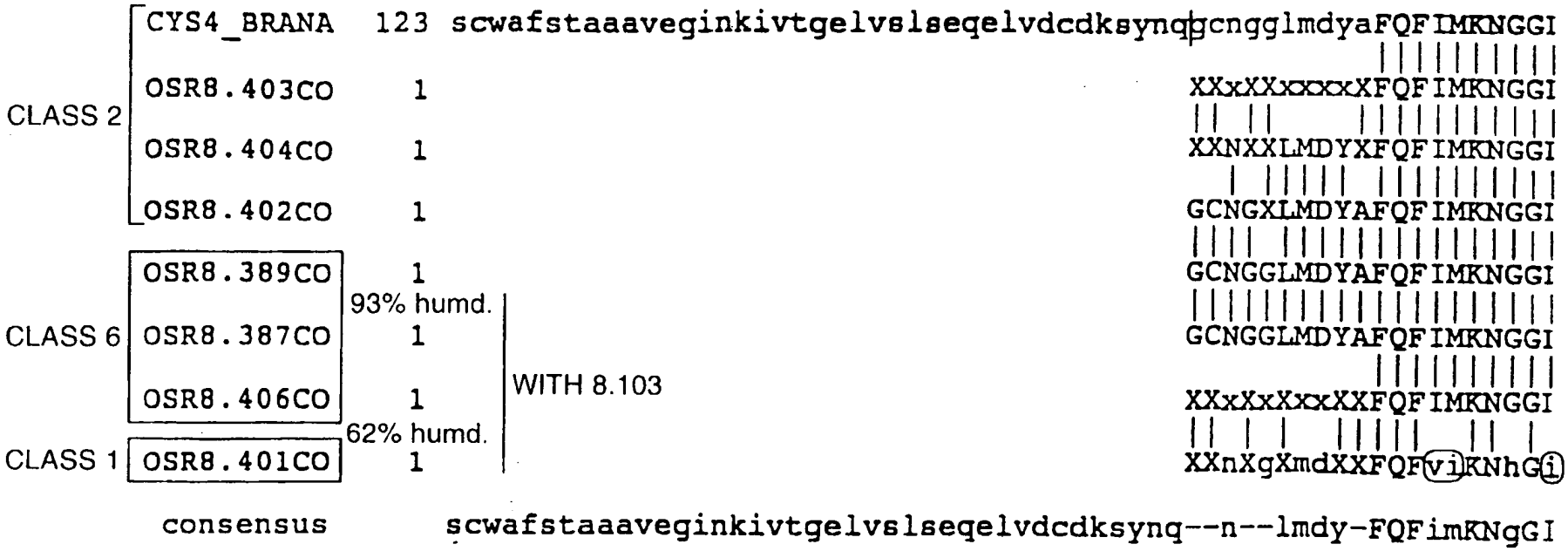


Fig.7. 2/4

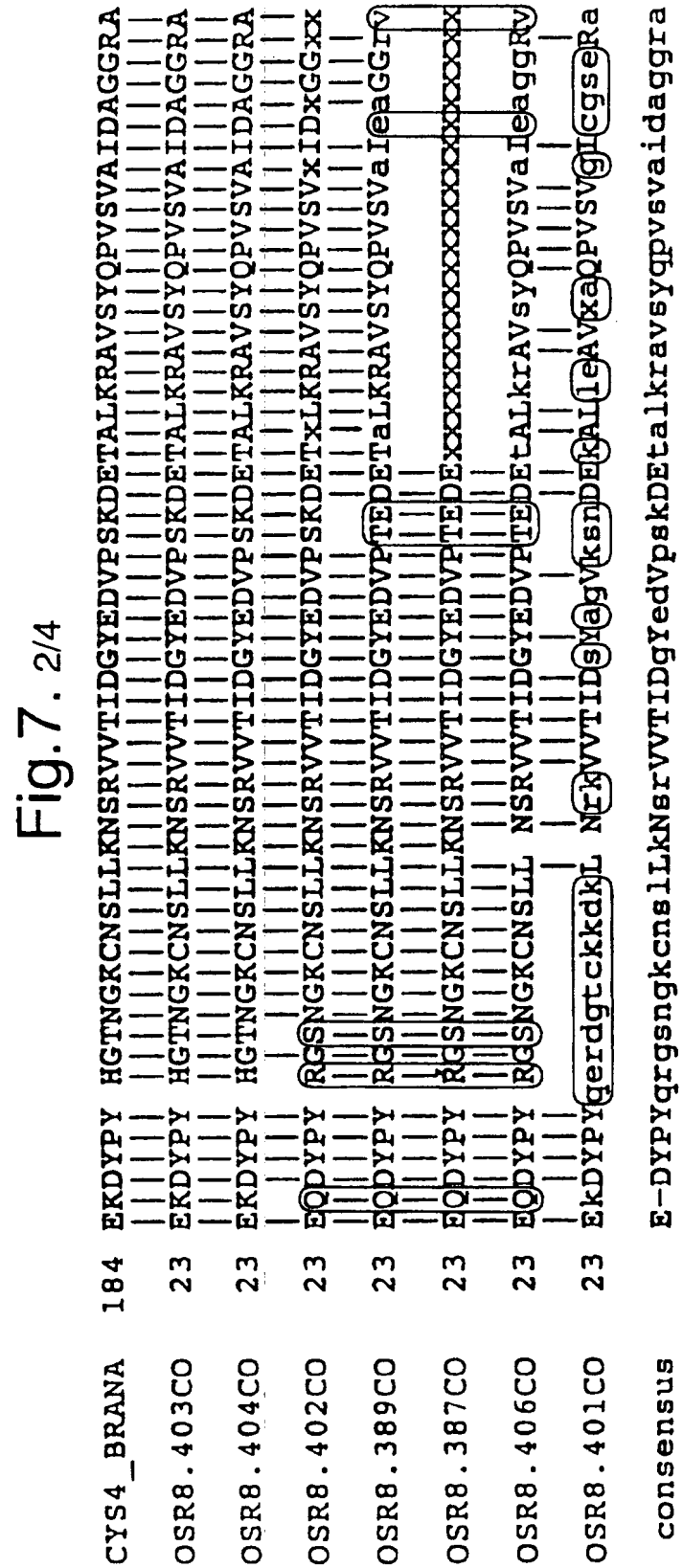

CYS4_BRANA 184 EKDYPY HGTNGKCNSLLKNSRVVTIDGYEDVPSKDETALKRAVSYQPVSVAIDAGGRA
OSR8.403CO 23 EKDYPY HGTNGKCNSLLKNSRVVTIDGYEDVPSKDETALKRAVSYQPVSVAIDAGGRA
OSR8.404CO 23 EKDYPY HGTNGKCNSLLKNSRVVTIDGYEDVPSKDETALKRAVSYQPVSVAIDAGGRA
OSR8.402CO 23 EQDYPY RGSNGKCNSLLKNSRVVTIDGYEDVPSKDETxLKRAVSYQPVSVxIDxGGxx
OSR8.389CO 23 EQDYPY RGSNGKCNSLLKNSRVVTIDGYEDVPTEDETaLKRAVSYQPVSVaIeaGGrV
OSR8.387CO 23 EQDYPY RGSNGKCNSLLKNSRVVTIDGYEDVPTEDExxxxxxxxxxxxxxxxxxxxxxx
OSR8.406CO 23 EQDYPY RGSNGKCNSLL NSRVVTIDGYEDVPTEDEtALkrAVsyQPVSVaIeaggRV
OSR8.401CO 23 EkDYPYgerdgtckkdkL NrkVVTIDsMagVksnDEkALleAVxaQPVSVgIcgseRa
consensus E-DYPYqrgsngkcnslLkNsrVVTIDgYedVpskDEtalkravsyqpvsvaidaggra

Fig.7. 3/4

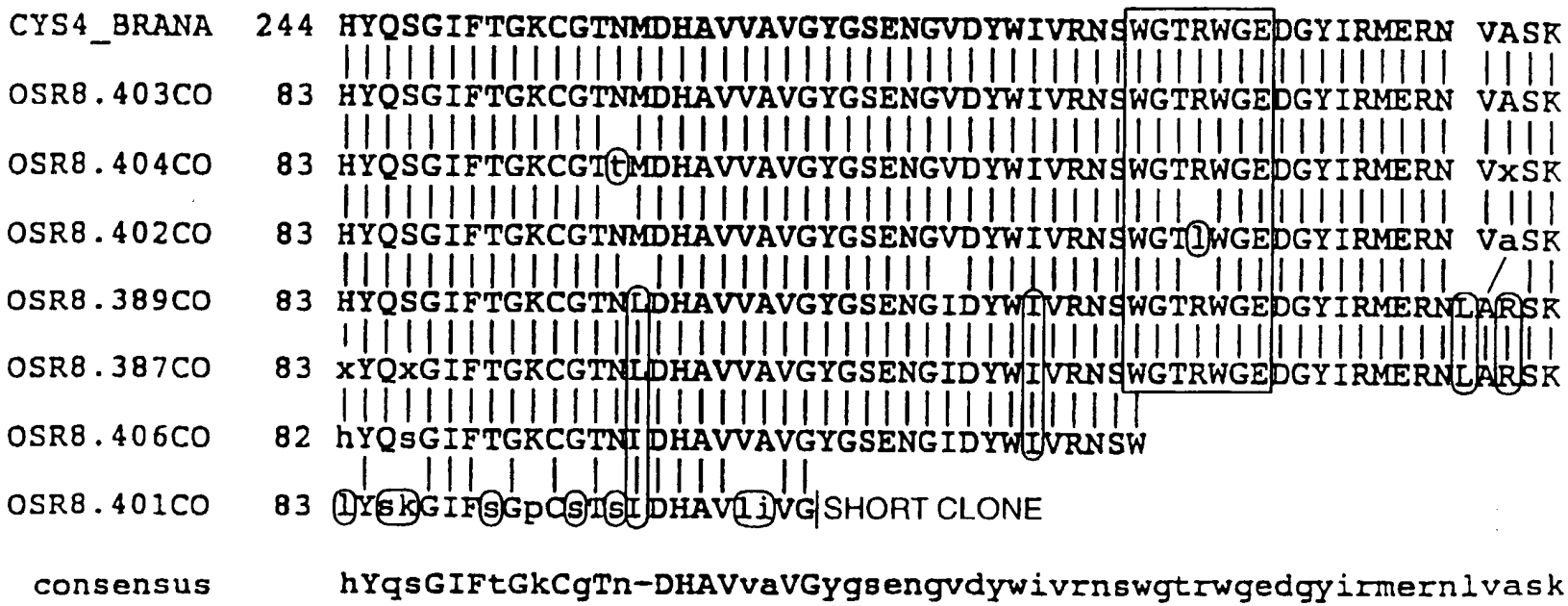


Fig.7. 4/4

```
CYS4_BRANA   304 KCGIaIEASYPVKYspnpvrgtssv|                    328
                 |||||||||||||||||||||||||
OSR8.403CO   143 KCGIxIEASYPVKY|SHORT CLONE
                 |||||||||||||||
OSR8.404CO   143 KCGIAIEASYPVKYSPNPVRgTSSV
                 |||||||||||||||||||||||||
OSR8.402CO   143 KCGIAIEASYPVKYSPNPVRxTSSV
                 |||||||||||||||||||||||||
OSR8.389CO   144 KCGIxVxxSYxVKYSPNPVRGTSSV
                 |||||||||||||||||||||||||
OSR8.387CO   144 KxxIaVeaSYpVKYSPNPxRGTSSV
                 |||||||||||||||||||||||||
OSR8.406CO   121
                 |||||||||||||||
OSR8.401CO   106
                 |||||||||||||||
consensus        kcgiaieasypvkyspnpvrgtssv|
                                          168
                                           ↑
Alignment score = 3484.00                STOP
```

Fig.8.

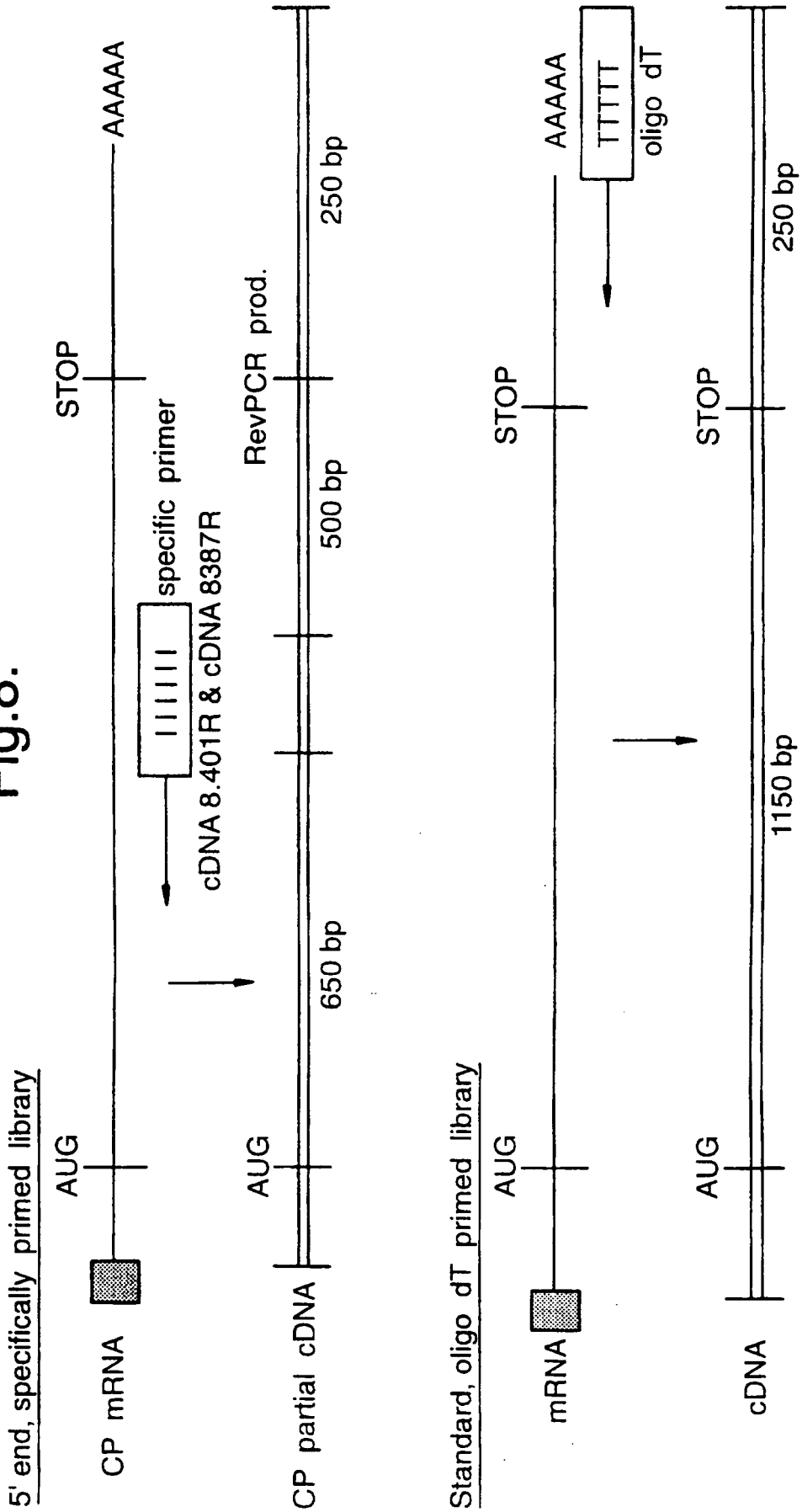

5' end, specifically primed library
CP mRNA
AUG
STOP
AAAAA
specific primer
cDNA 8.401R & cDNA 8387R
CP partial cDNA
AUG
650 bp
RevPCR prod.
500 bp
250 bp
Standard, oligo dT primed library
mRNA
AUG
STOP
AAAAA
TTTTT
oligo dT
cDNA
AUG
STOP
1150 bp
250 bp

Fig.9. 1/2

```
AMINO-Res-length    = 2
DELetion-weight     = 1.00
LEngth-factor       = 0
Matching-weight     = 5.00
NUCLEIC-Res-length  = 4
SPread-factor       = 90

Clustered order of selected sequences:

19. CYS4_BRANA =COT 44        (1-328)
 9. CYS2UP6_1  }              (1-225)
11. CYS2UP7_1  }CLASS 2       (1-230)
10. CYS2UP8_2  }              (1-225)

Region Alignment: (listed in Clustered order)

CYS4_BRANA     1          ATG                                                  mSIYLR
                           |                                                    |||||
 CYS2UP6_1     1       QTYNMASSPKLLSLLLLYVFVSLASGYESIISDNHLSLPSDRSWRTDEEVISIYLR
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CYS2UP7_1     1 rkptKQTYNMASSPKLLSLLLLYVFVSLASGYESIISDNHLSLPSDRSWRTDEEVISIYLR
                     |||||||||  ||||||||||  ||||  ||||  ||||  ||||||||||||  ||||
 CYS2UP8_2     1      KQTYNMASSaKLLSLLLLYVFiSLASsdESIInDNHLiLPSDRSWRTDEEVmSIYLk

 consensus       rkptkqtynmasspkllslllyvfvslasgyesiisdnhlslpsdrswrtdeev-SIYLr
```

Fig.9. 2/2

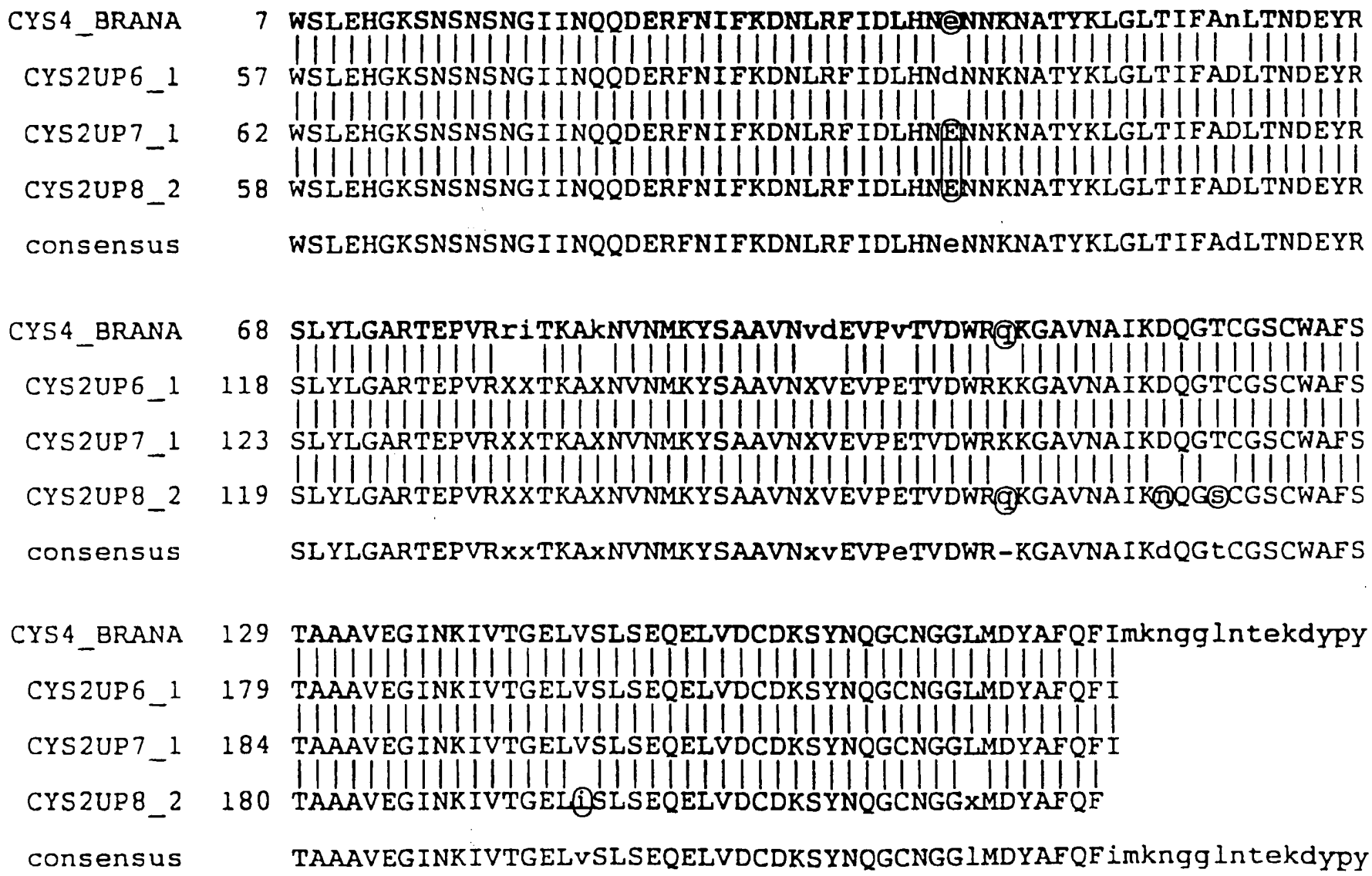


Fig.10. 1/6

Sequencing results

```
AMINO-Res-length    = 2
DELetion-weight     = 1.00
LEngth-factor       = 0
Matching-weight     = 5.00
NUCLEIC-Res-length  = 4
SPread-factor       = 90

Clustered order of selected sequences:

   9. COT44          (1-1102)
  10. CYS2UP6        (1-675)
  11. CYS2UP7        (1-691)
  12. CYS2UP8        (1-679)
   3. CYS6UP3NCOD    (1-322)
   1. CYS6UP5NCOD    (1-307)
   4. CYS6UP2NCOD    (1-288)
   2. CYS6UP4NCOD    (1-255)
```

Fig.10. 2/6

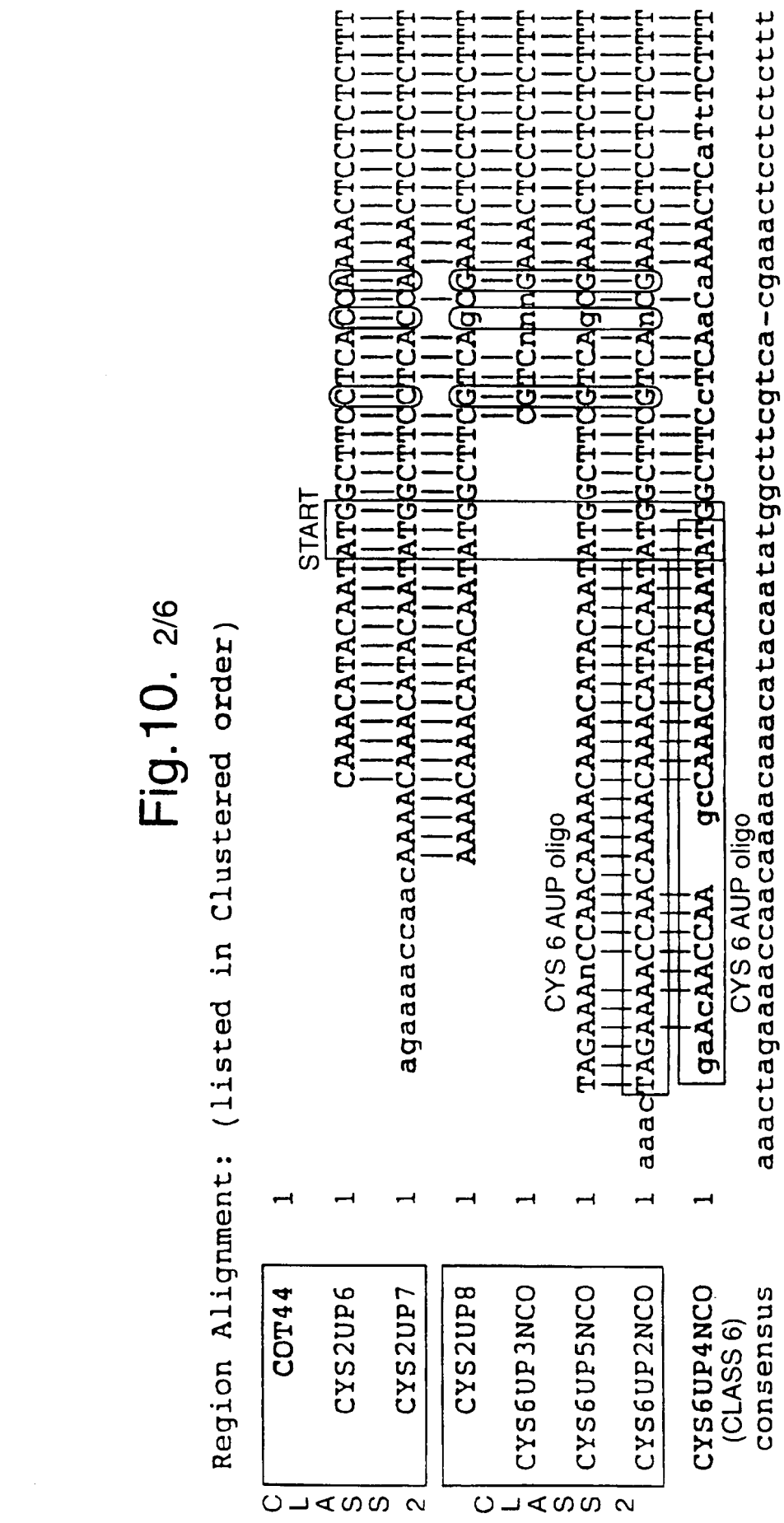

Region Alignment: (listed in Clustered order)
CLASS 2
CLASS 2
COT44
CYS2UP6
CYS2UP7
CYS2UP8
CYS6UP3NCO
CYS6UP5NCO
CYS6UP2NCO
CYS6UP4NCO
(CLASS 6)
consensus
START
CYS 6 AUP oligo
CYS 6 AUP oligo

Fig.10. 3/6

```
COT44          1
CYS2UP6       42 ACTTCTCTTATACGTCTTCGTTTCGTTAGCCTCCGGTTATGAGTCCATCATCAGTGACAAC
CYS2UP7       57 ACTTCTCTTATACGTCTTCGTTTCGTTAGCCTCCGGTTATGAGTCCATCATCAGTGACAAC
CYS2UP8       46 ACTTCTCTTGTACGTCTTCATTTCATTAGCCTCCaGTGATGAGTCCATCATCAACGACAAC
CYS6UP3NCO    23 ACTTCTCTTGTACGTCTTCATTTCATTAGCCTCCnGTGATGAGTCCATCATCAACGACAAC
CYS6UP5NCO    58 ACTTCTCTTGTACGTCTTCATTTCATTAGCCTCCAGTGATGAGTCCATCATCAACGACAAC
CYS6UP2NCO    62 ACTTCTCTTGTACGTCTTCATTTCATTAGCCTCCAGTGATGAGTCCATCATCAACGACAAC
CYS6UP4NCO    54 ACTTCTCcTaTACGTCgTCgTTTCATTAGCCTCCgGTGATGAGTCCActAcCAttaACAAC
consensus        acttctcttgtacgtcttcatttcattagcctcc-gtgatgagtccatcatcaacgacaac
```

Fig.10. 4/6

```
COT44        1                                         tGATG           TCCATCTACT
CYS2UP6    103  CATCTCAGTCTTCCATCTGACCGTTCGTGGAGAACCGATGAAGAAGTGATATCCATCTACT
CYS2UP7    118  CATCTCAGTCTTCCATCTGACCGTTCGTGGAGAACCGATGAAGAAGTGATATCCATCTACT
CYS2UP8    107  CATCTCATTCTTCCATCTGACCGCTCGTGGAGAACCGATGAAGAAGTGATGTCCATCTACT
CYS6UP3NCO  84  CATCTCATTCTTCCATCTGACCGCTCGTGGAGAACCGATGAAGAAGTGATGTCCATCTACT
CYS6UP5NCO 119  CATCTCATTCTTCCATCTGACCGCTCGTGGAGAACCGATGAAGAAGTGATGTCCATCTACT
CYS6UP2NCO 123  CATCTCATTCTTCCATCTGACCGCTCGTGGAGAACCGATGAAGAAGTGATGTCCATCTACT
CYS6UP4NCO 115  CATCTCAaTCTTCCATCgGACnGCTCaTGGAGAACCGATGAAGAAGTGAgGTCCATCTACT

consensus       catctcattcttccatctgaccgctcgtggagaaccGATGaagaagtgatgTCCATCTACT
```

Fig.10. 5/6

```
COT44      16  TAAGATGGTCCTTGGAGCACGGGAAAAGTAACAGCAACAGCGGTATTATCAACCAACA
CYS2UP6   164  TAAGATGGTCCTTGGAGCACGGGAAAAGTAACAGCAACAGCGGTATTATCAACCAACA
CYS2UP7   179  TAAGATGGTCCTTGGAGCACGGGAAAAG@AACAGCAACAGCGGTATTATCAACCAACA
CYS2UP8   168  TAAAATGGTCCTTGGAGCACGGGAAAAGTAACAGCAACAGCGGTATTATCAACCAACA
CYS6UP3NCO 145 TAAAATGGTCCTTGGAGCACGGGAAAAGTAACAGCAACAGCGGTATTATCAACCAACA
CYS6UP5NCO 180 TAAAATGGTCCTTGGAGCACGGGAAAAGTAACAGCAACAGCGGTATTATCAACCAACA
CYS6UP2NCO 184 TAAAATGGTCCTTGGAGCACGGGAAAAGTAACAGCAACAGCGGTATTATCAACCAACA
CYS6UP4NCO 176 TAcAgTGGT   gtncGgAnGGGAAAA   ctAGCAACAaCAACGGTATcgTCAACCAACA
consensus      TAaaaTGGTccttggaGcAcGGGAAAAgtaacAGCAACAgCAACGGTATtaTCAACCAACA
```

Fig.10. 6/6

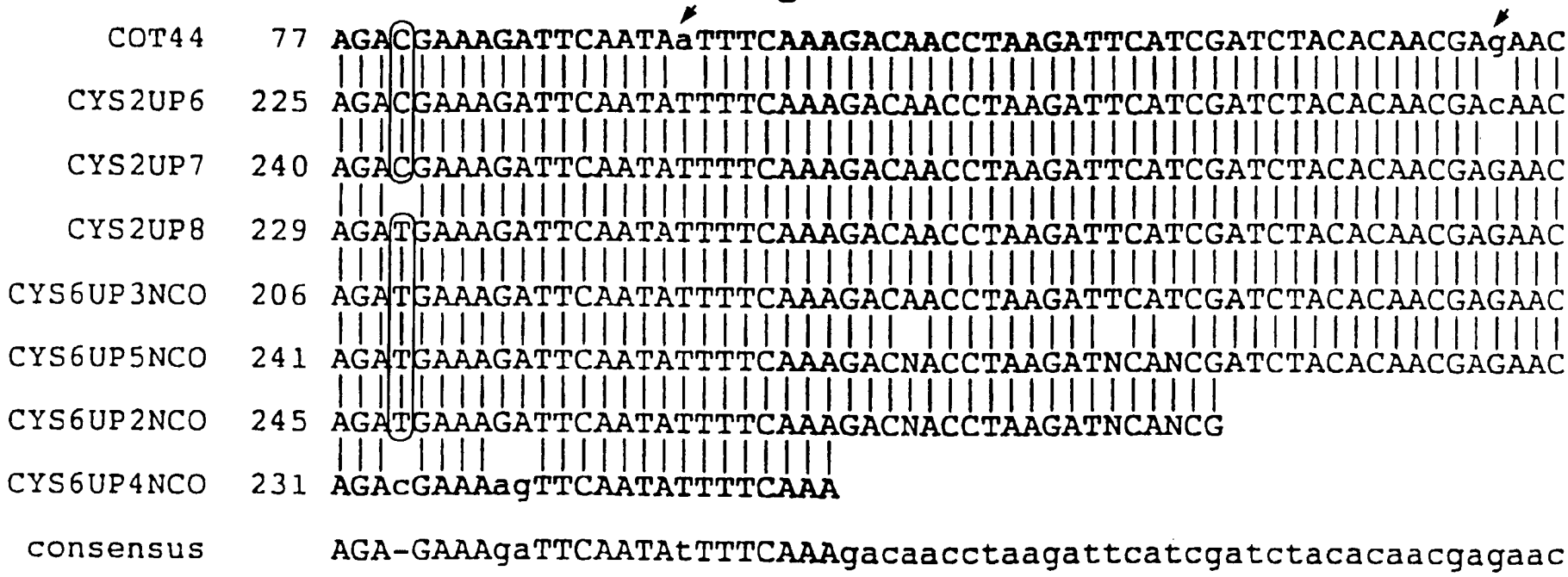


Fig.11. 1/21

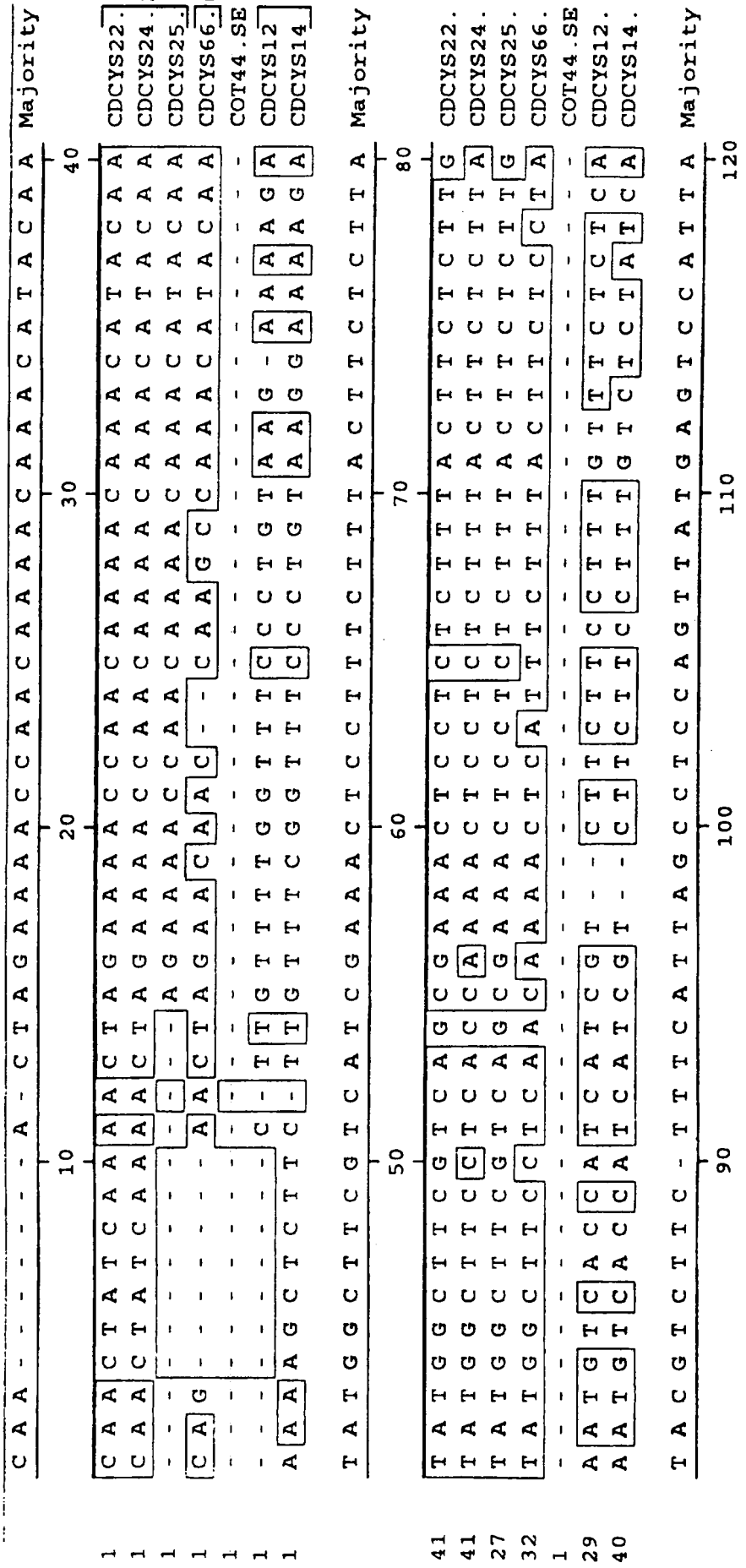

Majority
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12
CDCYS14
Majority
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority

Fig.11. 2/21

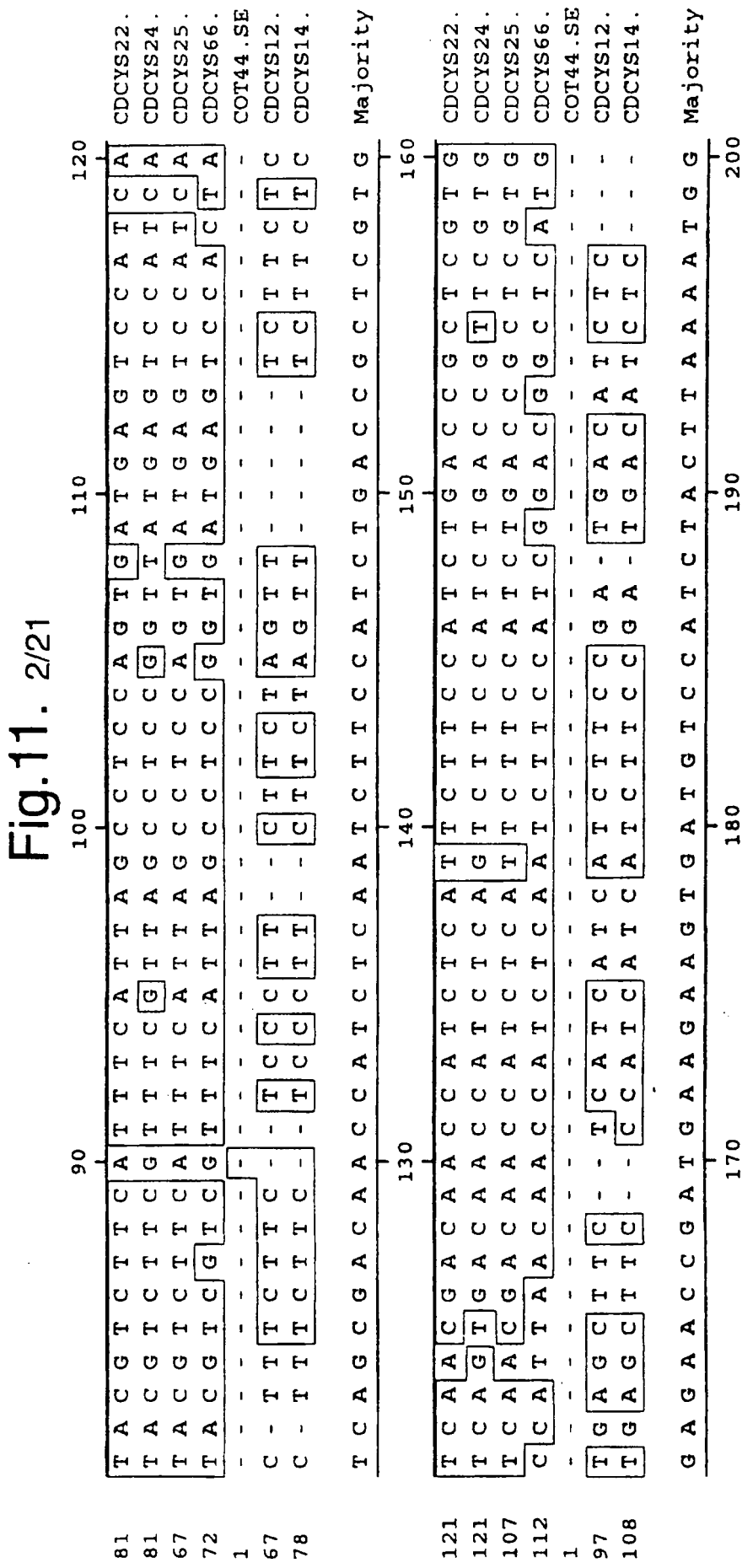

90
100
110
120
130
140
150
160
170
180
190
200
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority

Fig.11. 3/21

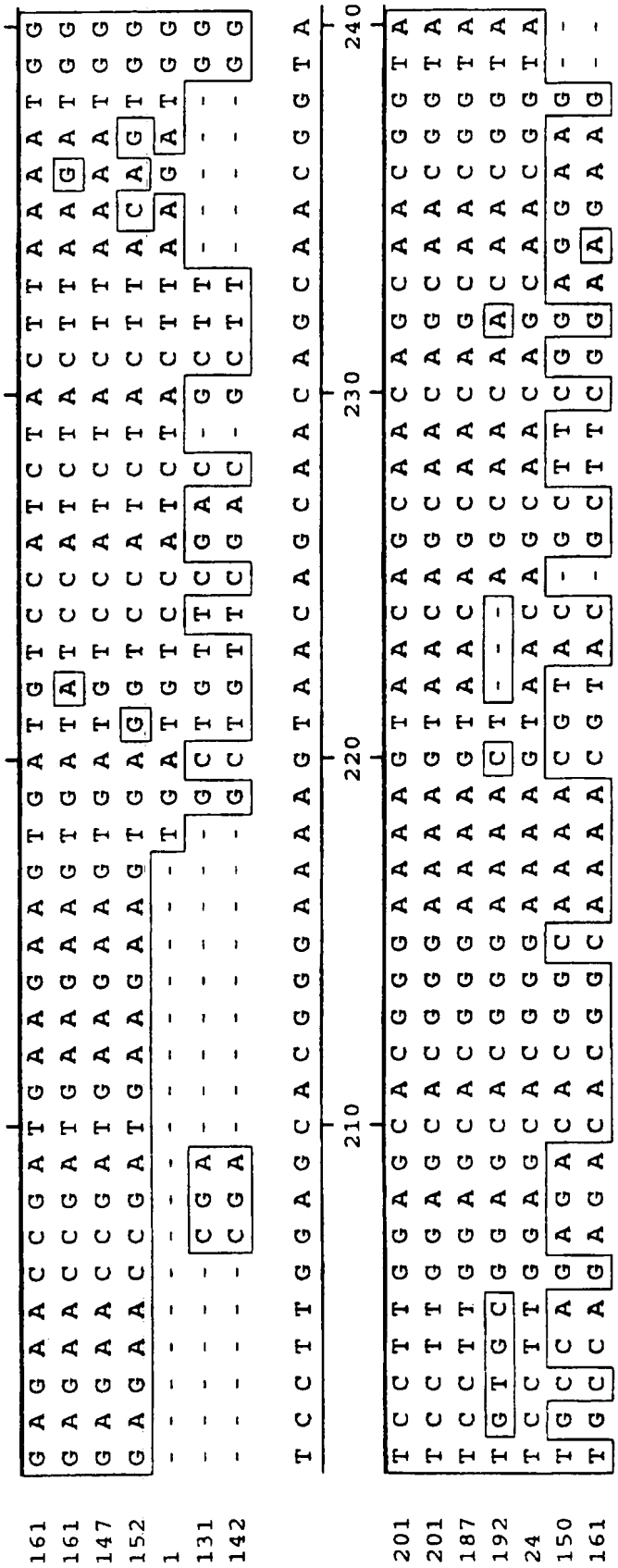

170
180
190
200
210
220
230
240
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority

Fig.11. 4/21

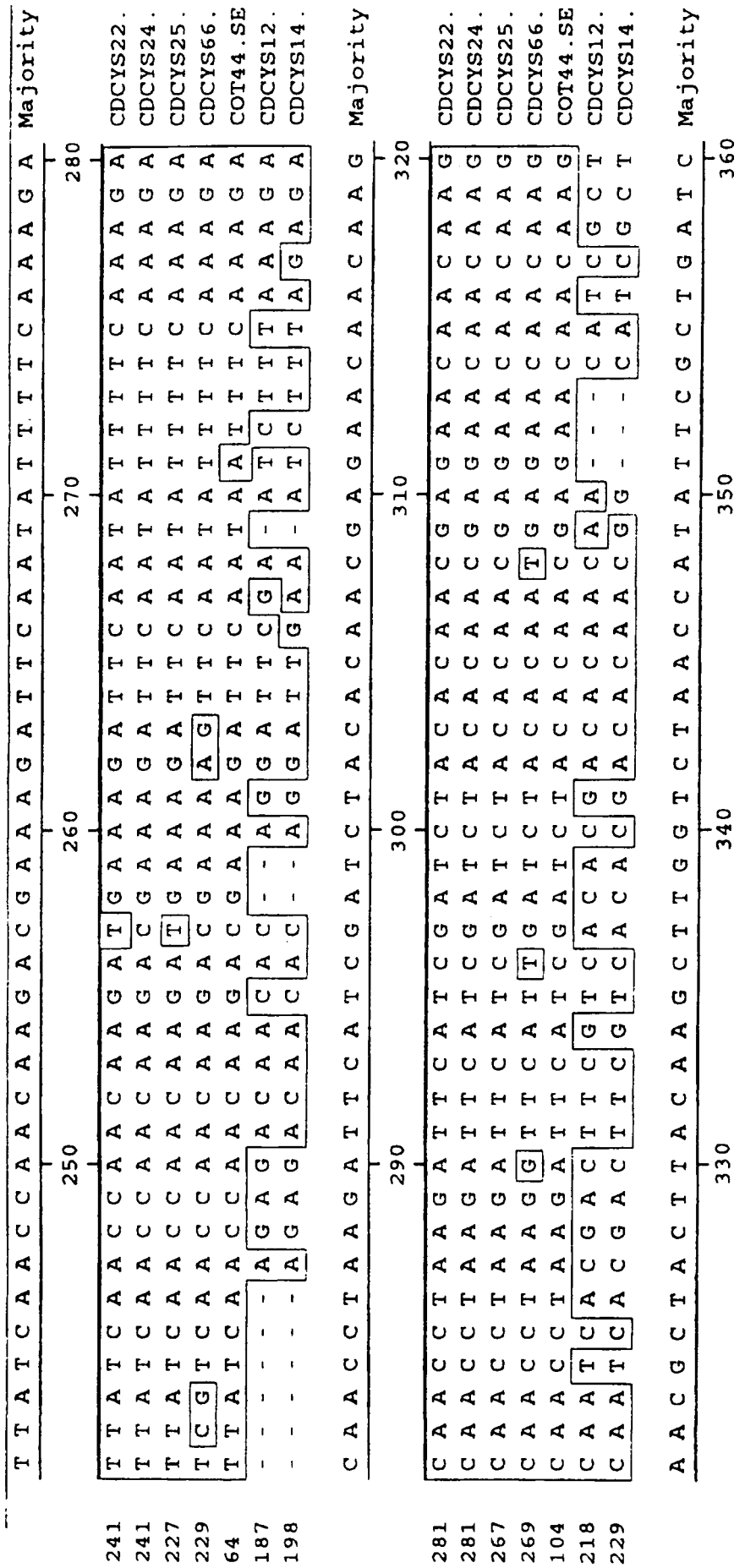

Majority
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
250
260
270
280
290
300
310
320
330
340
350
360

Fig.11. 5/21

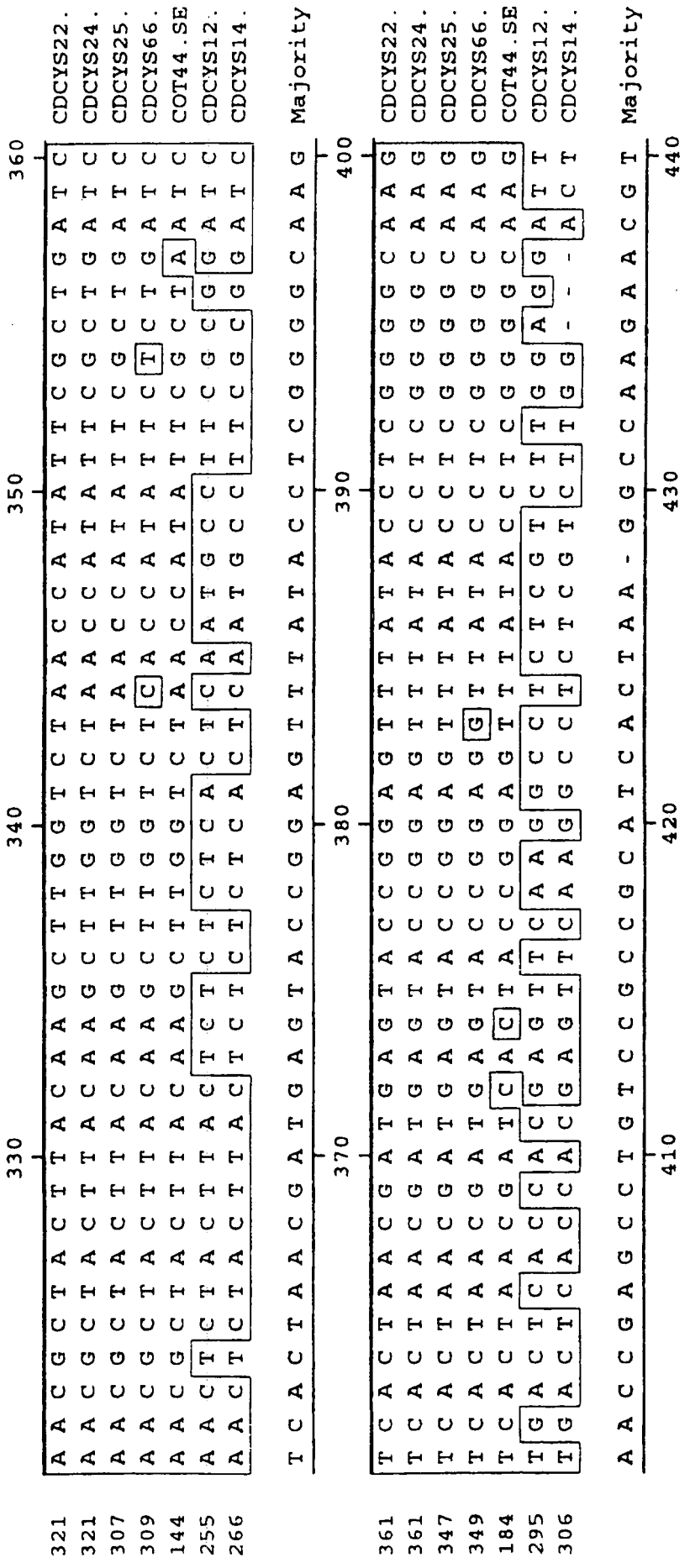

330
340
350
360
370
380
390
400
410
420
430
440
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority
321
321
307
309
144
255
266
361
361
347
349
184
295
306

Fig.11. 6/21

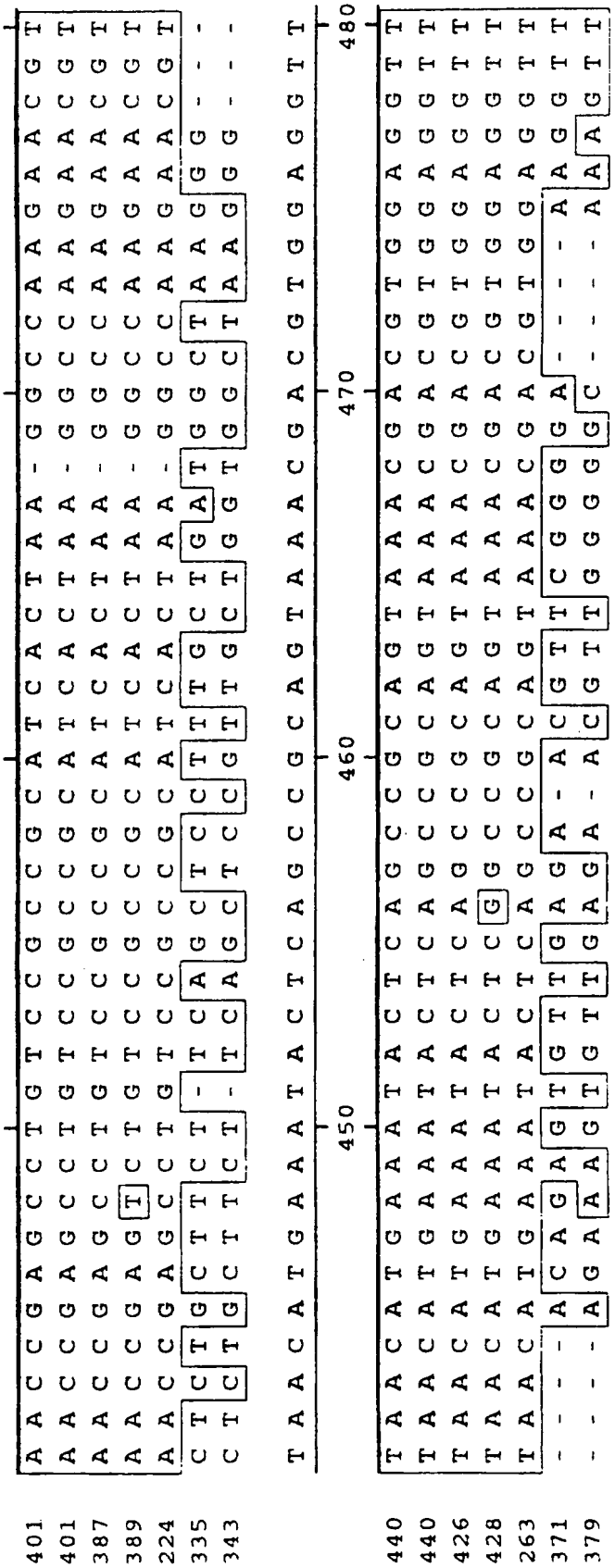
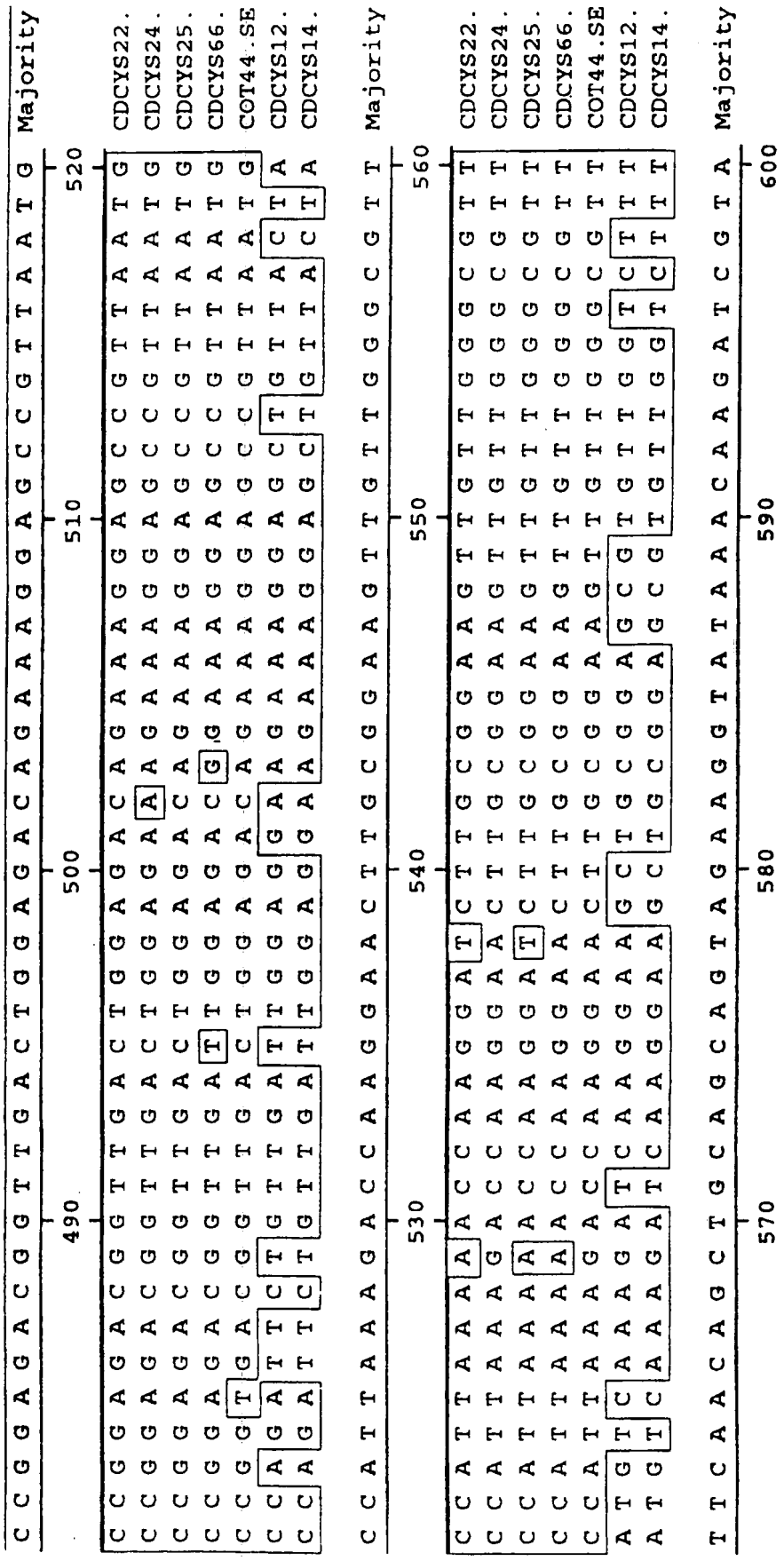

410
420
430
440
450
460
470
480
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority
401
401
387
389
224
335
343
440
440
426
428
263
371
379

Fig.11. 7/21


Majority
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
480
480
466
468
303
402
410
520
520
506
508
343
442
450
490
500
510
520
530
540
550
560
570
580
590
600

40/89

Fig.11. 8/21

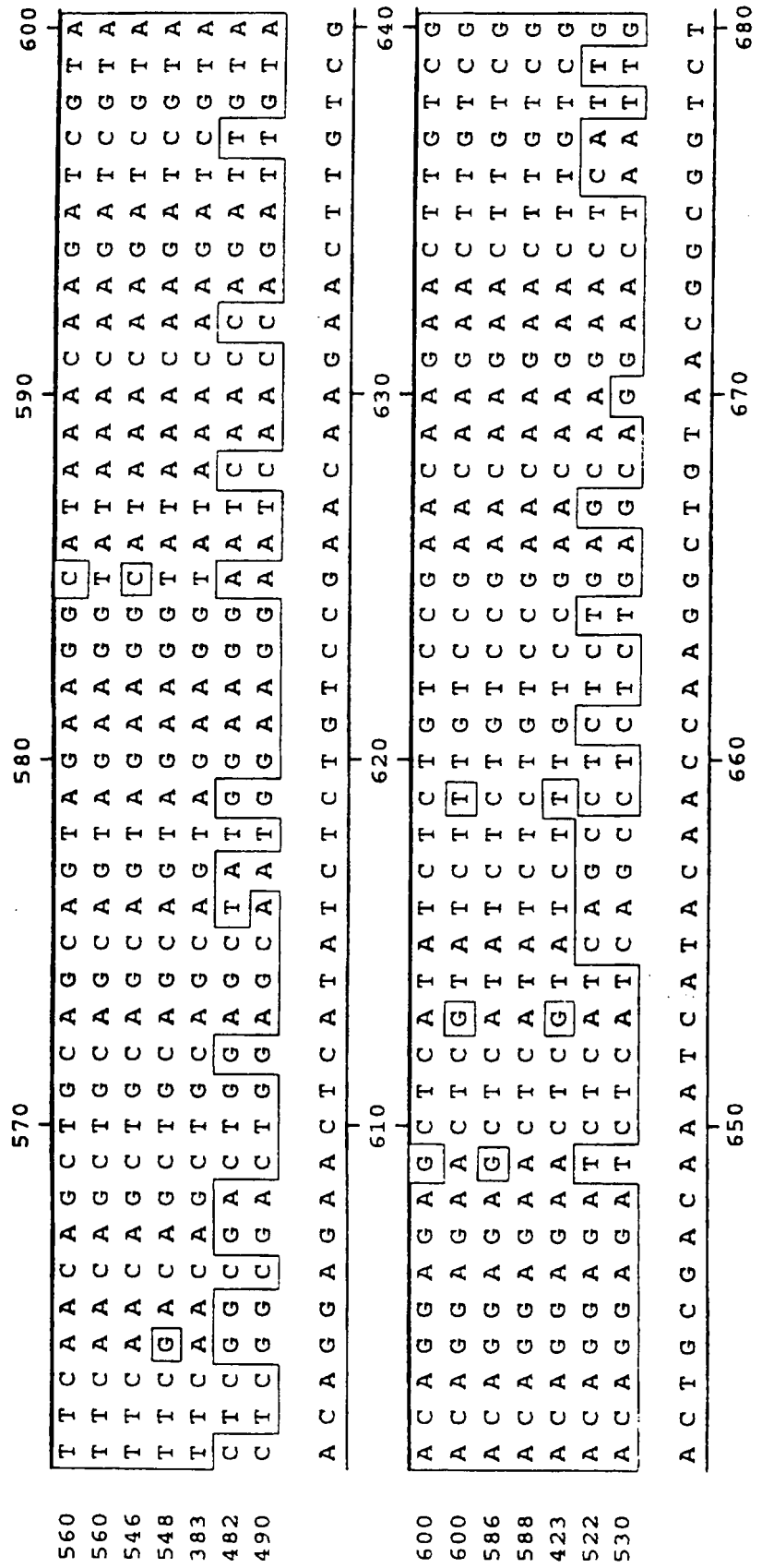


Fig.11. 9/21

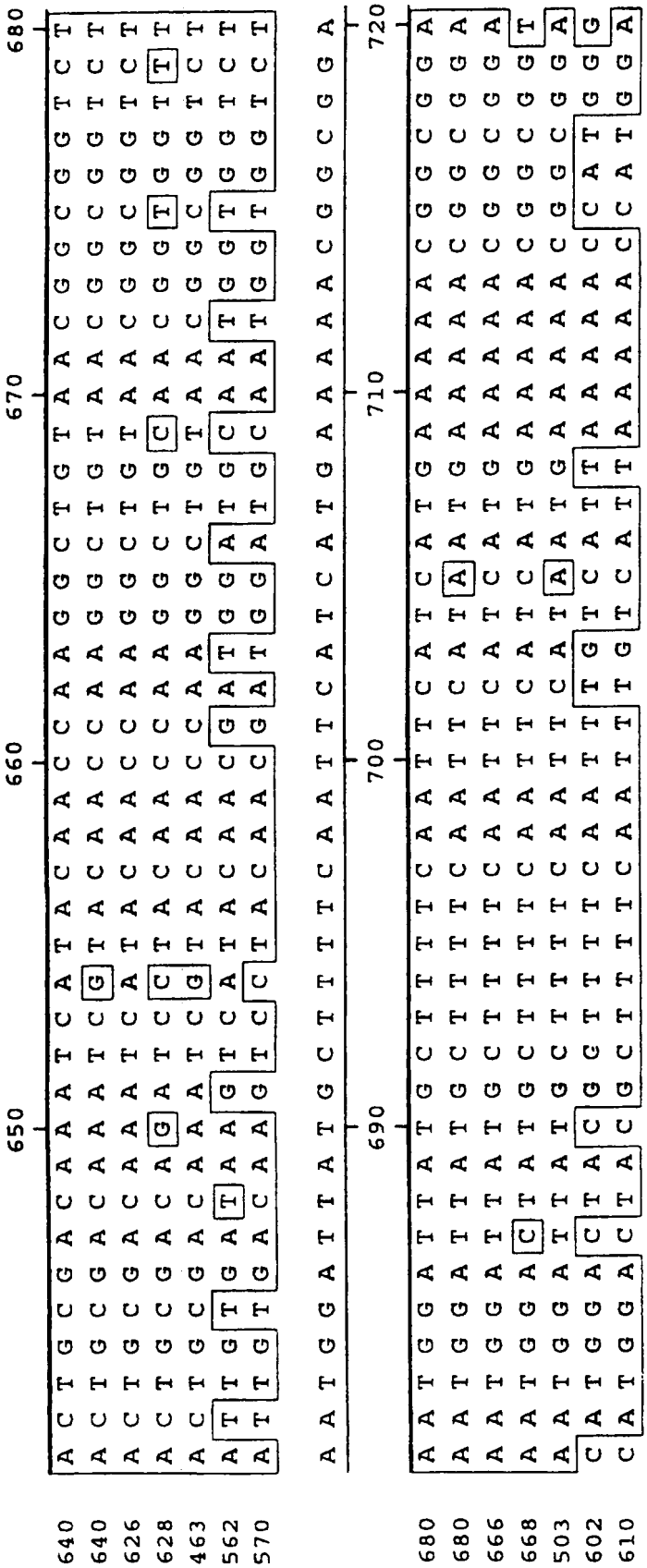

650
660
670
680
690
700
710
720
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority
640
640
626
628
463
562
570
680
680
666
668
503
602
610

Fig.11. 10/21

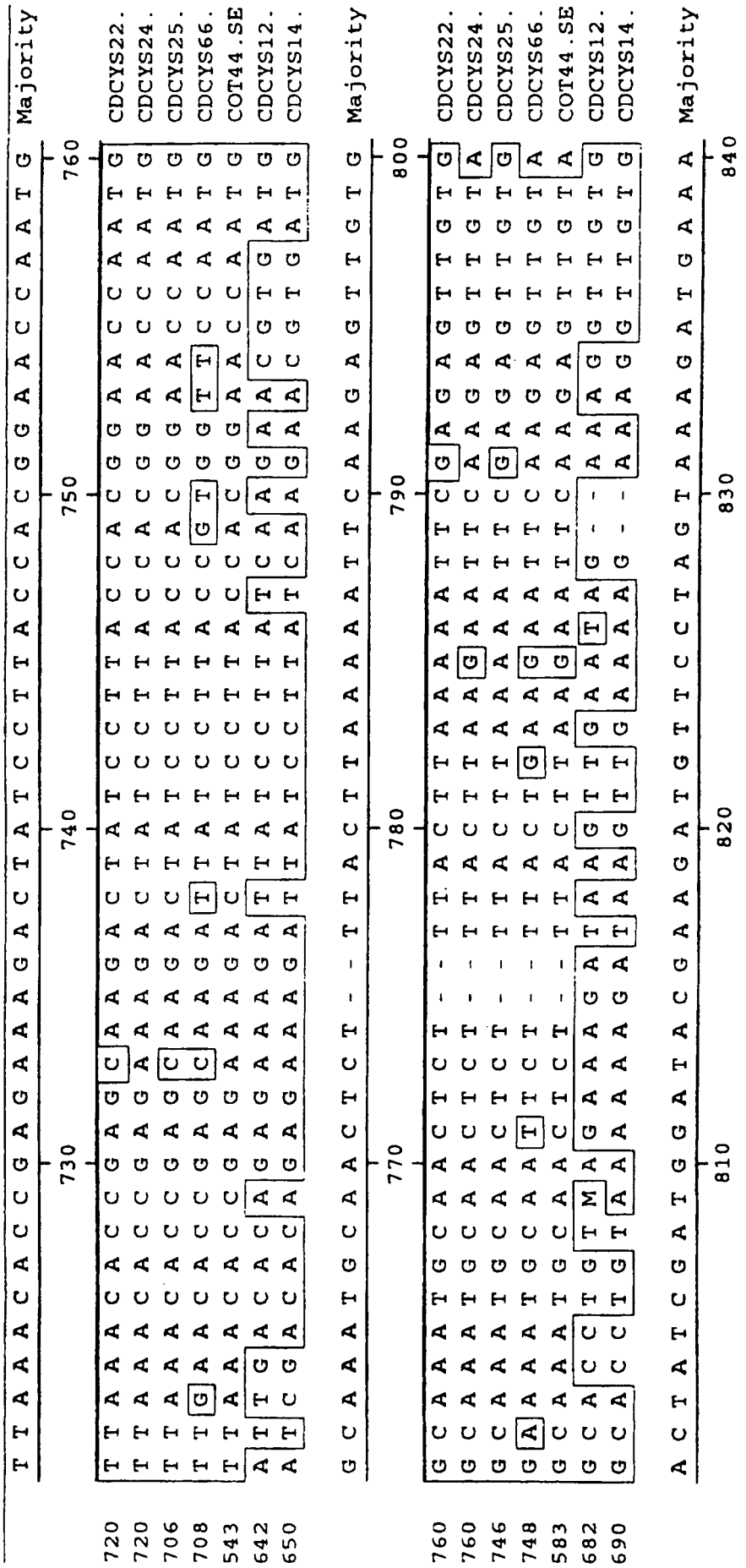

Majority
730
740
750
760
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
720
720
706
708
543
642
650
Majority
770
780
790
800
760
760
746
748
583
682
690
Majority
810
820
830
840

Fig.11. 11/21

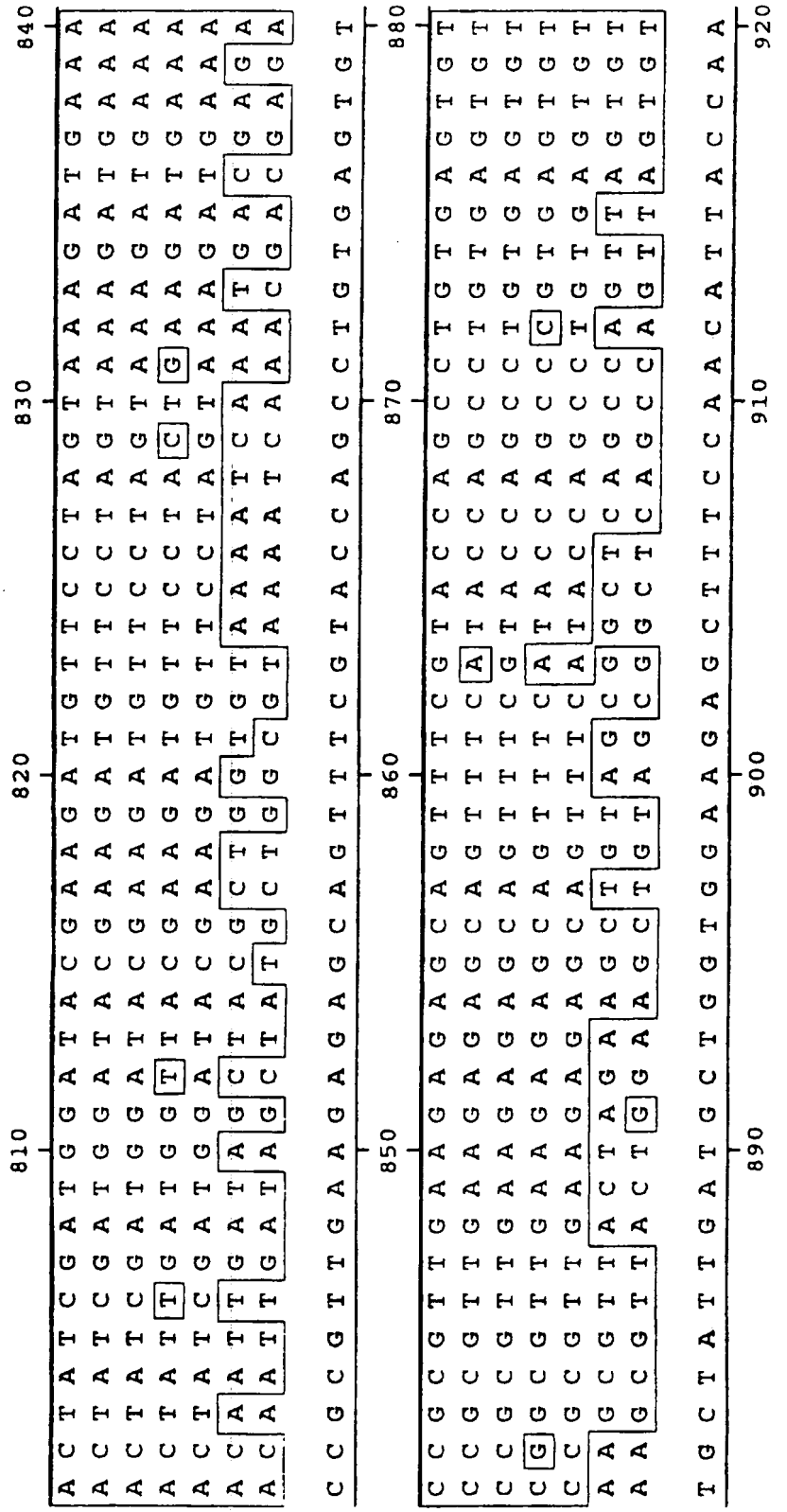

CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority
810
820
830
840
850
860
870
880
890
900
910
920
798
798
784
786
621
720
728
838
838
824
826
661
760
768

Fig.11. 12/21

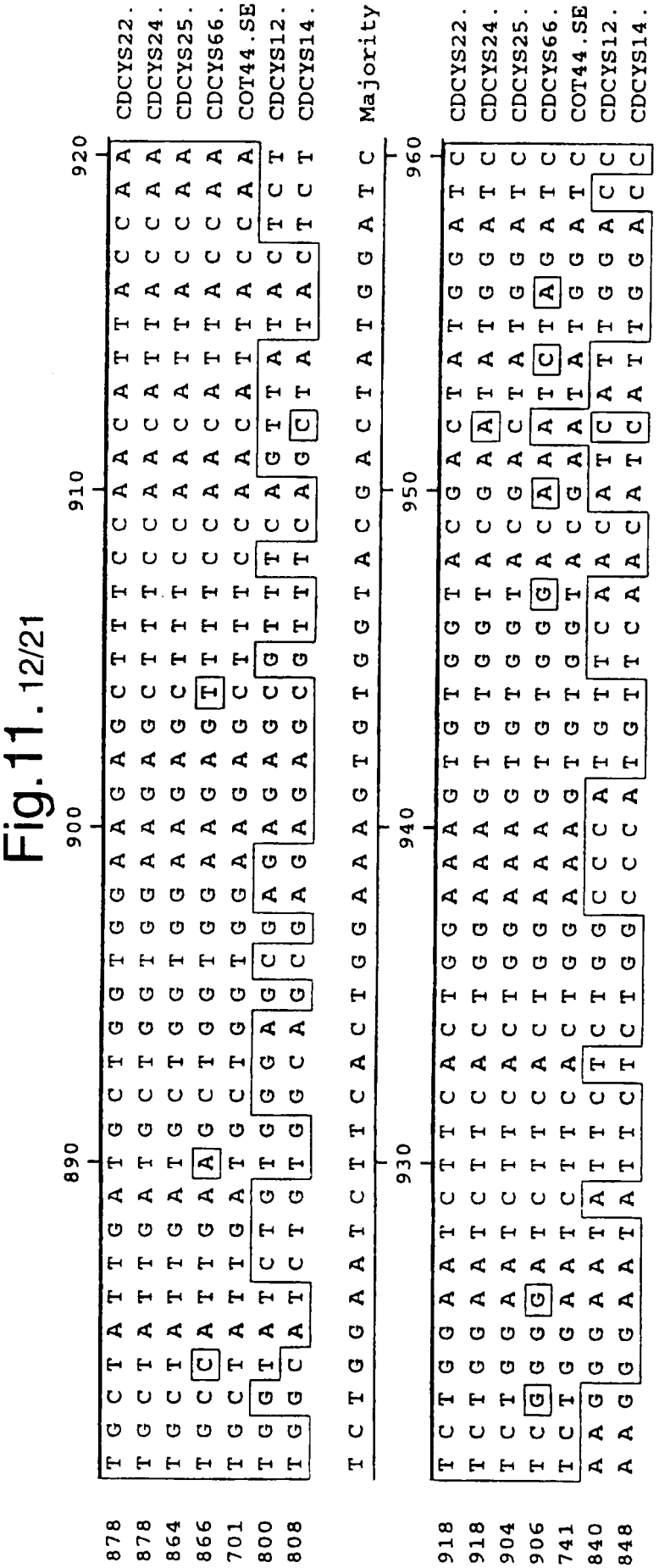

890
900
910
920
930
940
950
960
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority
878
878
864
866
701
800
808
918
918
904
906
741
840
848

Fig.11. 13/21

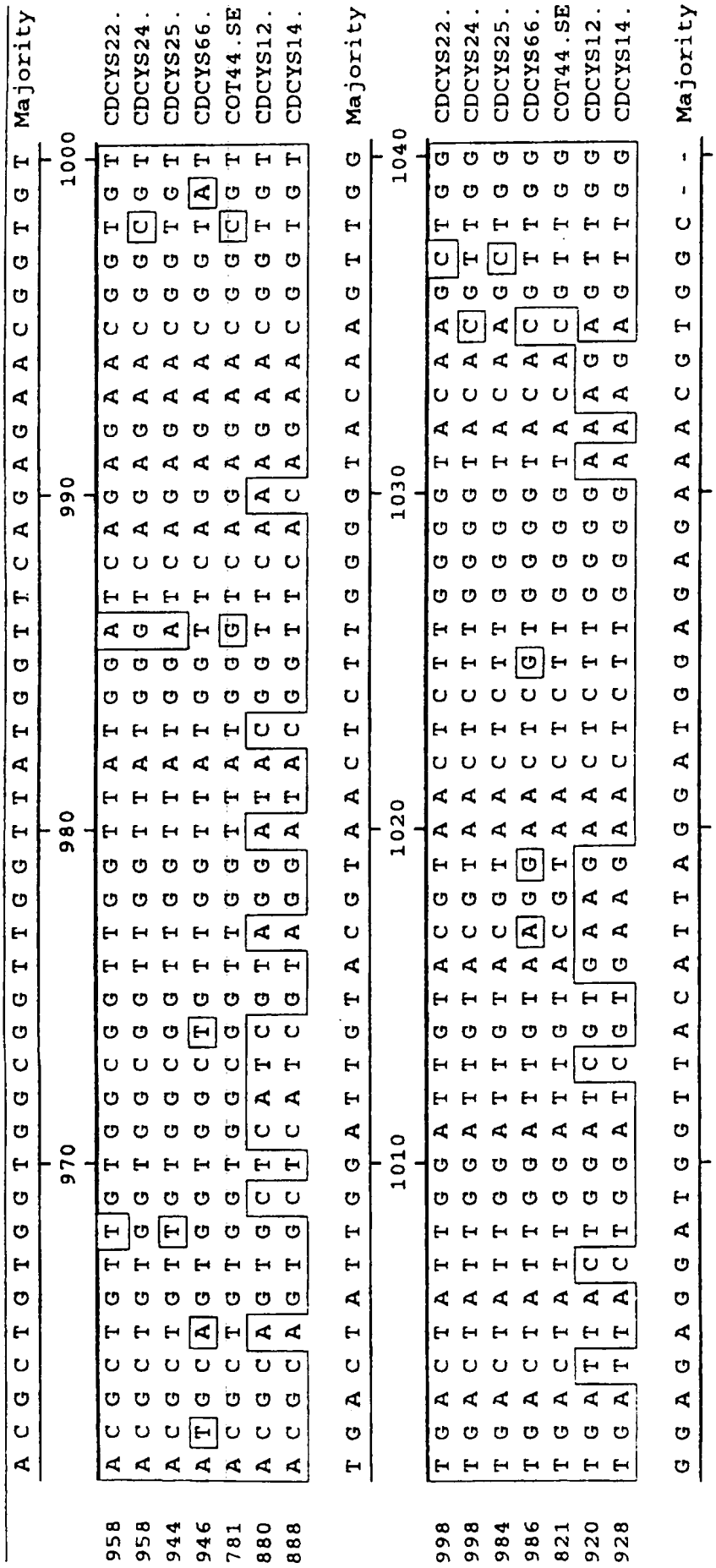

Majority
970
980
990
1000
1010
1020
1030
1040
1050
1060
1070
1080
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
958
958
944
946
781
880
888
998
998
984
986
821
920
928

Fig.11. 14/21

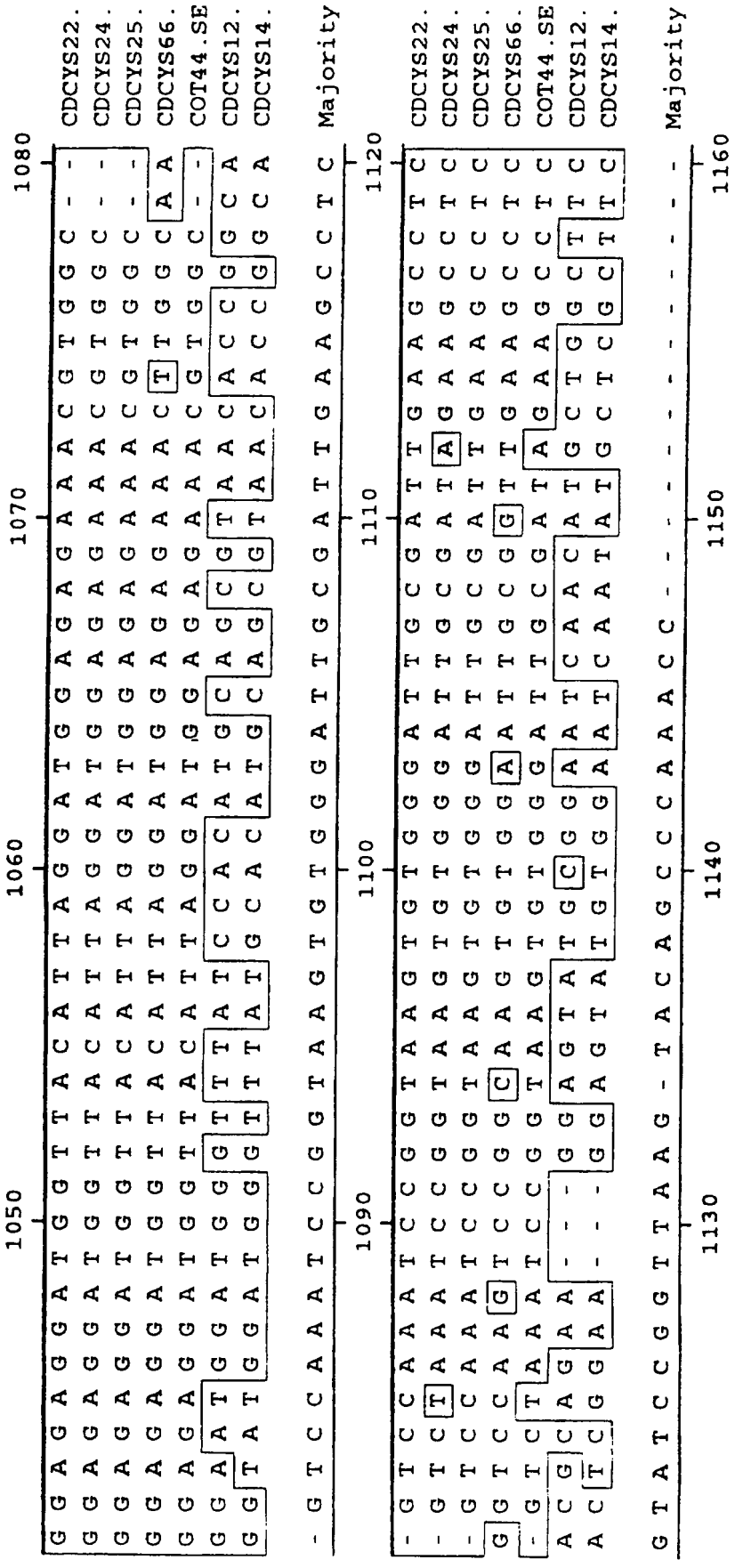

1050
1060
1070
1080
1090
1100
1110
1120
1130
1140
1150
1160
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority

Fig.11. 15/21

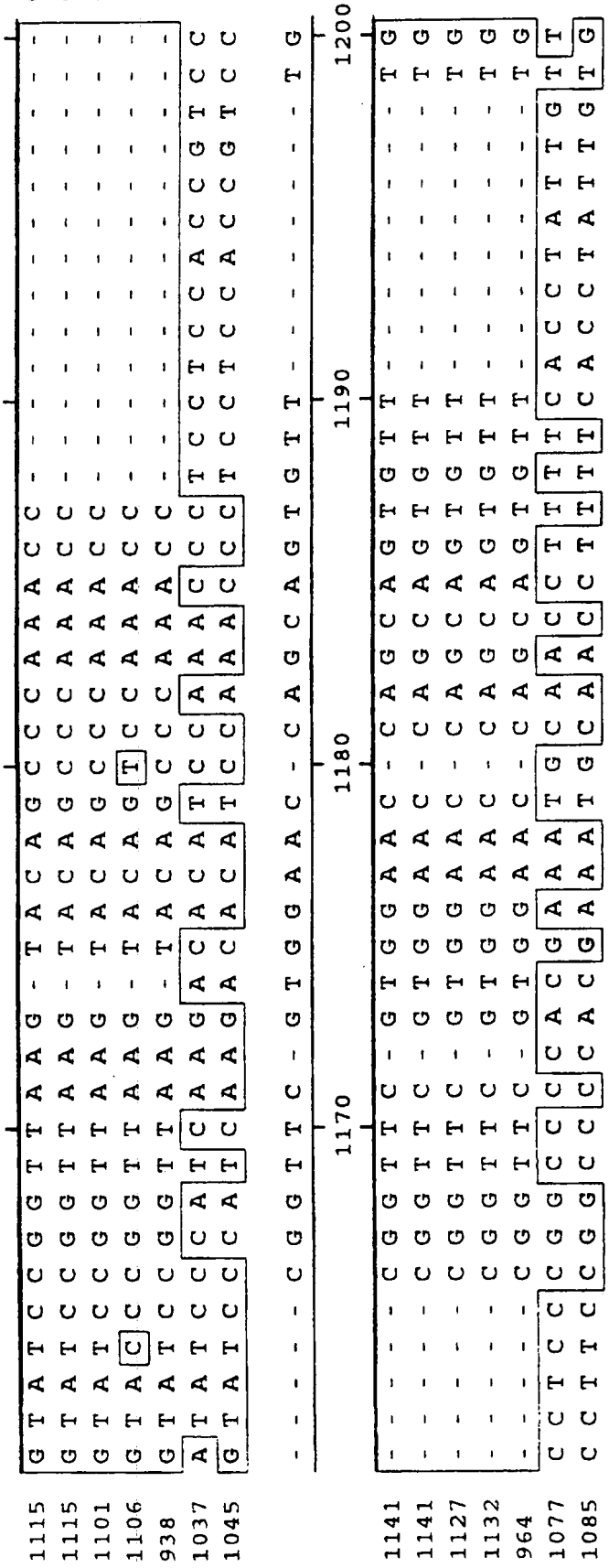

1130
1140
1150
1160
1170
1180
1190
1200
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority

Fig.11. 16/21

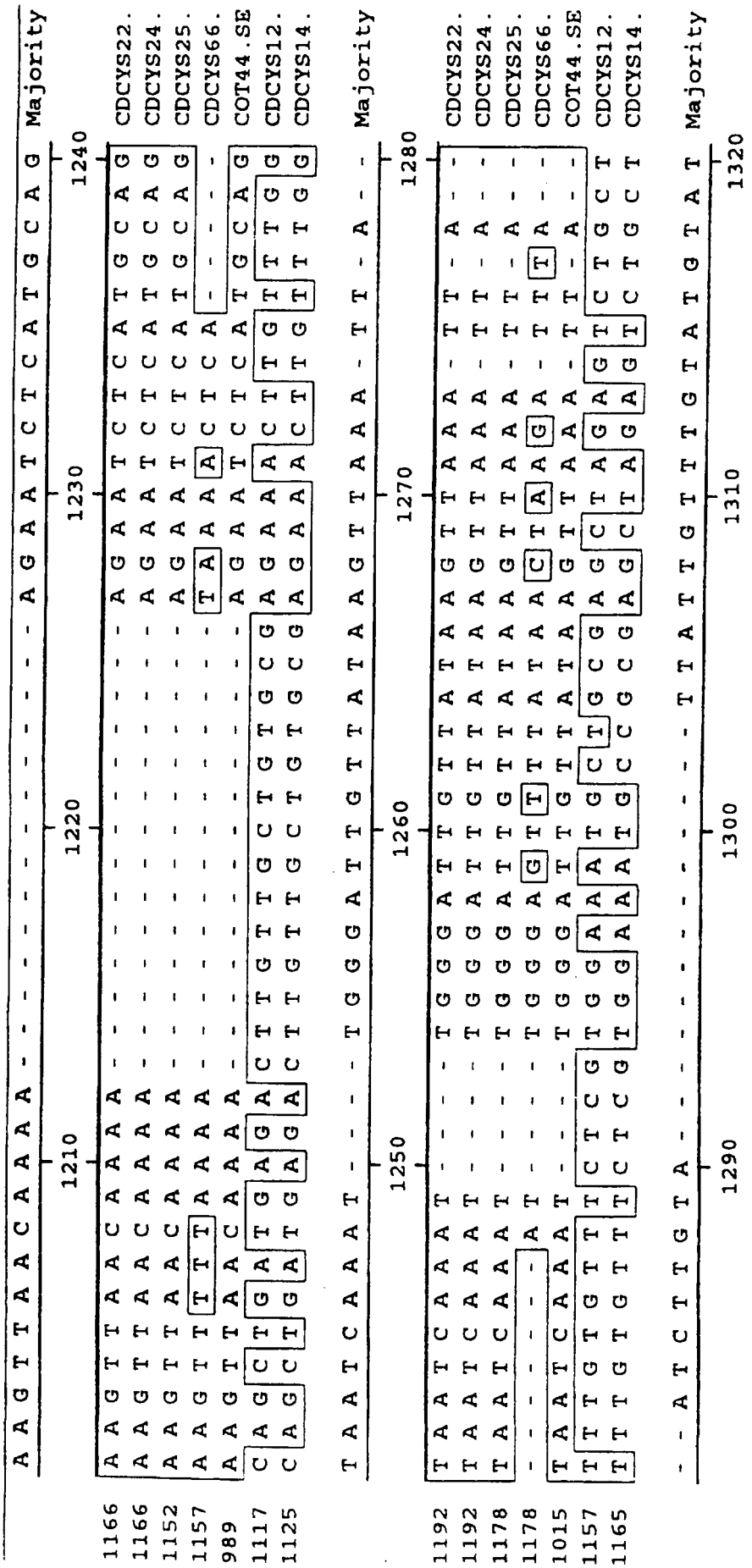

Majority
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
1210
1220
1230
1240
1250
1260
1270
1280
1290
1300
1310
1320

Fig.11. 17/21

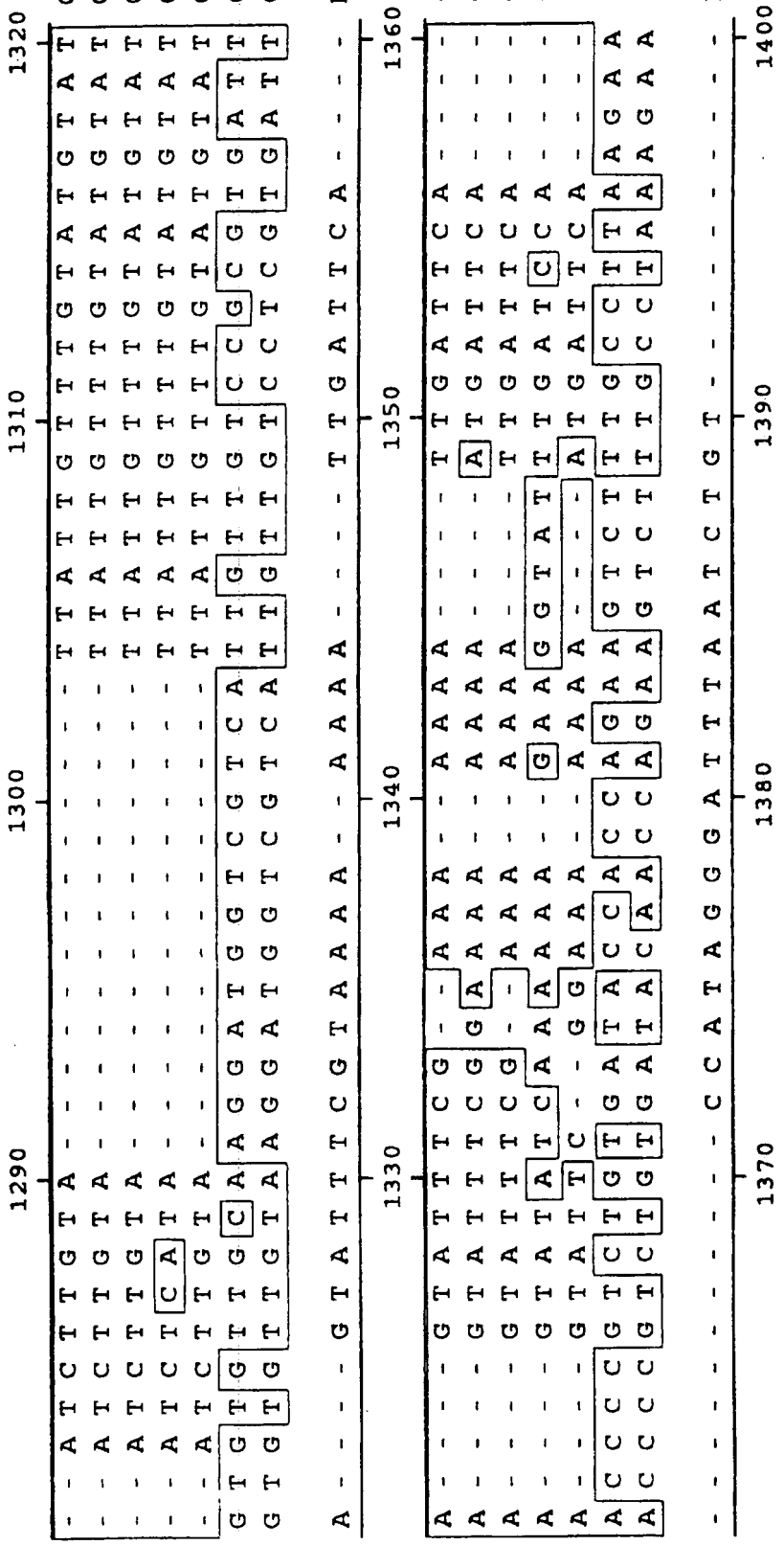

1290
1300
1310
1320
1330
1340
1350
1360
1370
1380
1390
1400
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority

Fig.11. 18/21

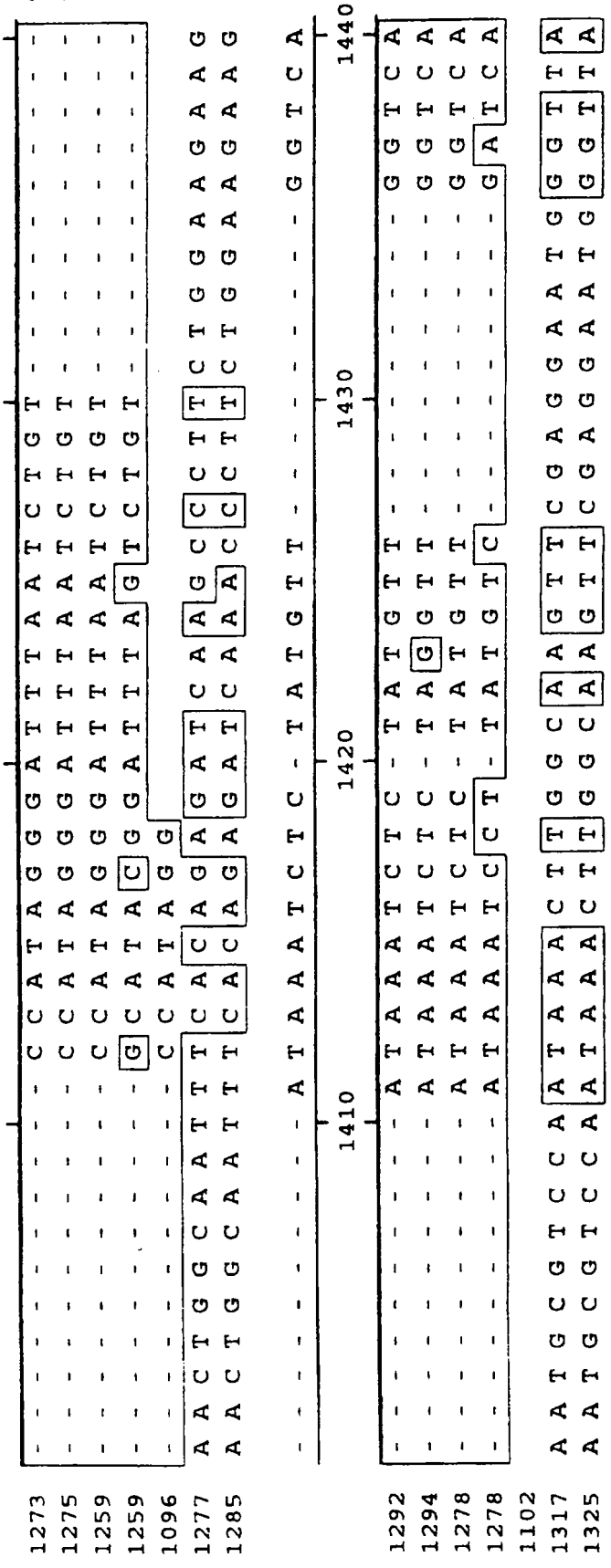

1370
1380
1390
1400
1410
1420
1430
1440
1273 CDCYS22.
1275 CDCYS24.
1259 CDCYS25.
1259 CDCYS66.
1096 COT44.SE
1277 CDCYS12.
1285 CDCYS14.
Majority
1292 CDCYS22.
1294 CDCYS24.
1278 CDCYS25.
1278 CDCYS66.
1102 COT44.SE
1317 CDCYS12.
1325 CDCYS14.

Fig.11. 19/21

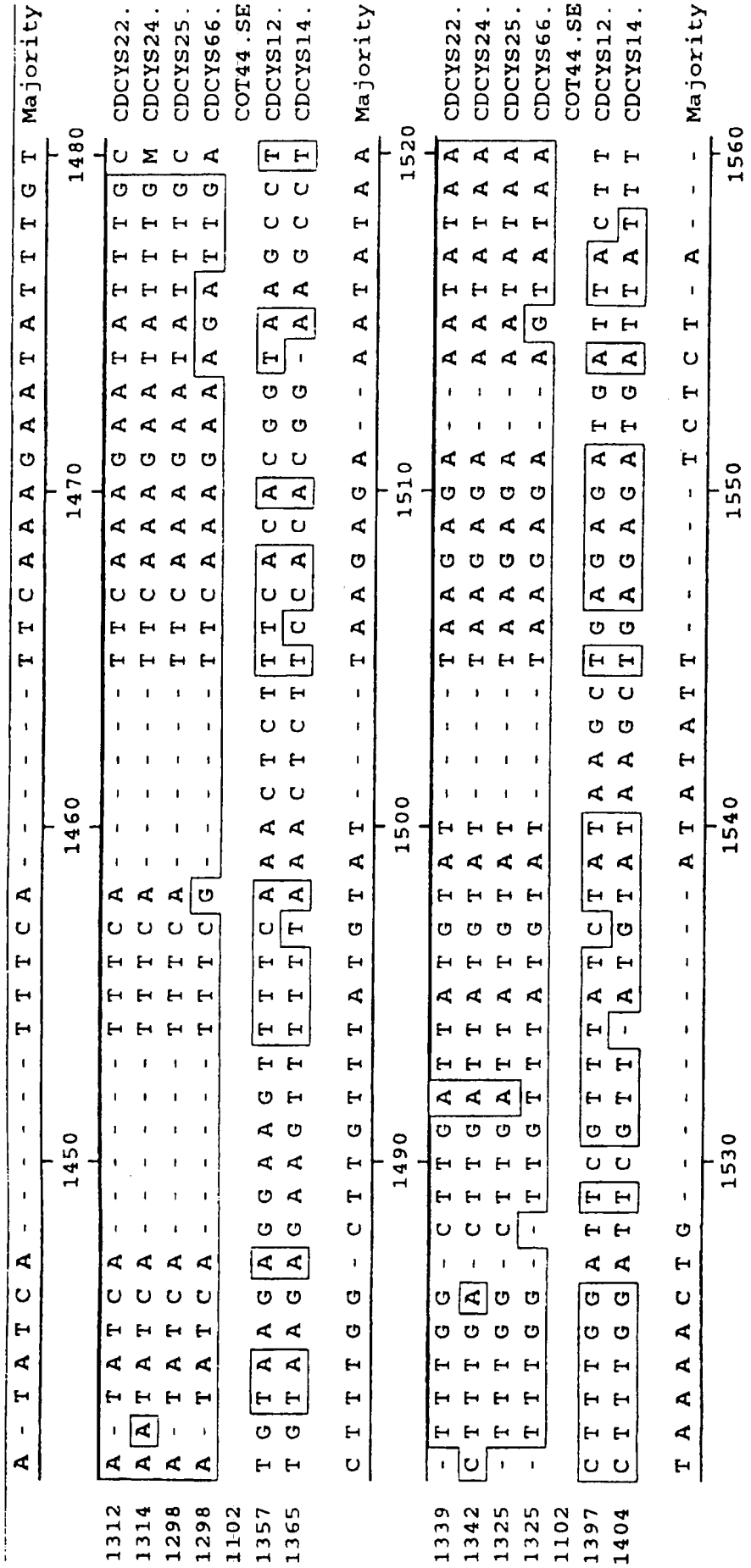

Majority
1450
1460
1470
1480
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
1490
1500
1510
1520
1530
1540
1550
1560

Fig.11. 20/21

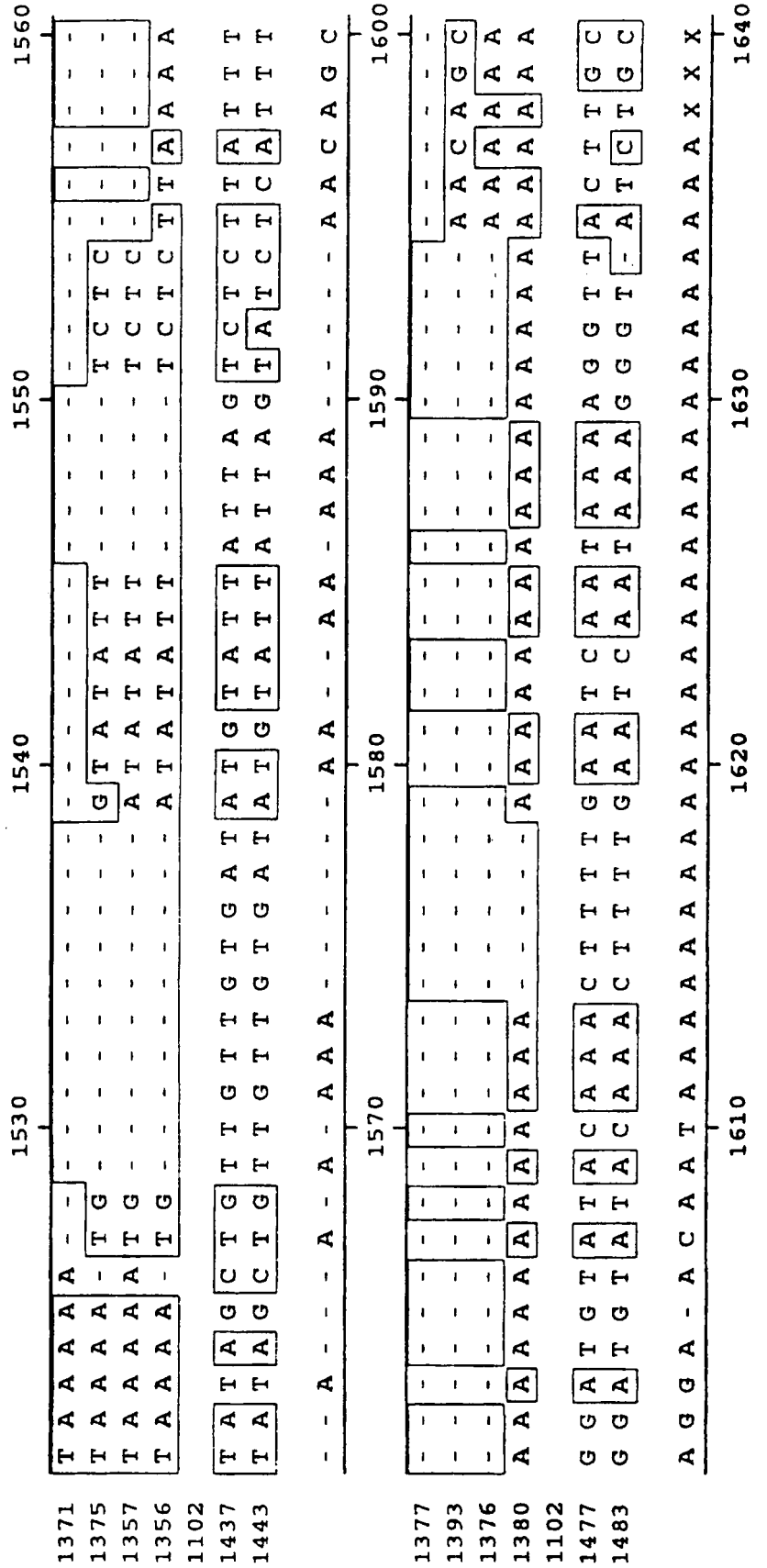
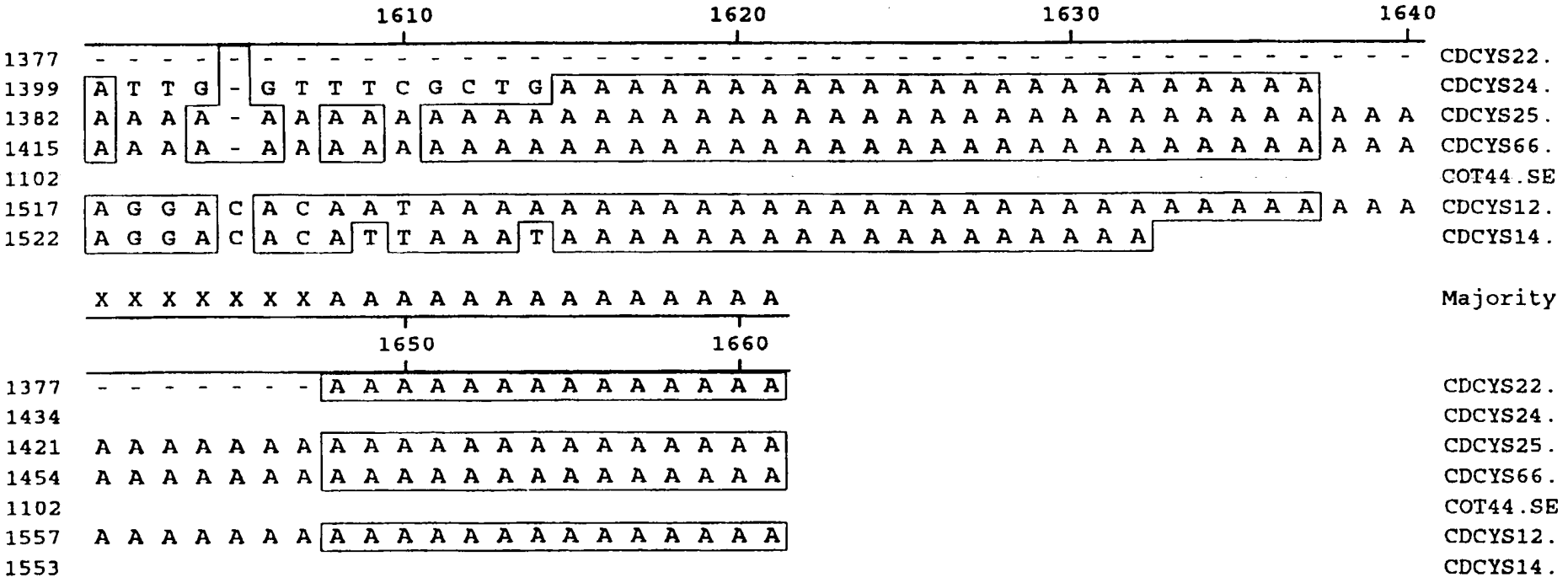

1530
1540
1550
1560
1570
1580
1590
1600
1610
1620
1630
1640
CDCYS22.
CDCYS24.
CDCYS25.
CDCYS66.
COT44.SE
CDCYS12.
CDCYS14.
Majority

Fig.11. 21/21



Decoration 'Decoration #1': Box residues that match the Consensus exactly.

Fig.12. 1/2

```
                    10        20        30        40        50        60
                    |         |         |         |         |         |
CTTGTTTTGG TTTCCCTGTA AGAAAAGAAA TGTCACCATC ATCGTCTTCT TCCTTTGTTT 60
CTCTCACTTT CTTCTCCCTT CTTCTAGTTT CTTCTCTGAG CTTCTCATCA TCATCTTCCG 120
ATGACATCTC CGAGCTGTTC GACGCTTGGT GCCAGAGACA CGGCAAAACG TACGCTTCGG 180
AGGAAGAGAG ACAACACAGG ATTCGAATCT TTAAAGACAA TCACGACTTC GTCACACGAC 240
ACAACAACAT CGCTAACTCT ACTTACTCTC TCTCACTCAA TGCCTTCGCG GATCTGACTC 300
ACCACGAGTT CAAGGCCTCT CGTCTTGGAG GATTCTCTGC TTCTTCAGCT CCTTTGCTGA 360
TGGCTAAGGG ACAGAGTGTT GAGAACGTTC GGGGAAAGGT TCCAGATTCT GTTGATTGGA 420
GGAAGAAAGG AGCTGTTACT AATGTCAAAG ATCAAGGAAG CTGCGGAGCG TGTTGGTCTT 480
TCTCGGCGAC TGGAGCTATG GAAGGAATCA ACCAGATTGT AACAGGAGAT CTCATCAGCC 540
TCTCTGAGCA AGAACTCATT GATTGTGATA AGTCATACAA CGATGGATGC AATGGTGGTC 600
TCATGGACTA CGCTTTTCAA TTTGTCATTA AAAACCATGG GATTGACACA GAGAAAGATT 660
ATCCTTATCA AGAACGTGAT GGCACCTGTM AGAAAGATAA GTTGAATAGA AAGGTTGTGA 720
CAATTGATAG CTACGCTGGT GTAAAATCAA ATGACGAGAA AGCGTTACTA GAAGCTGTAG 780
```

Fig.12. 2/2

```
CGGCTCAGCC AGTTAGTGTT GGTATCTGTG GGAGCGAGAG AGCGTTTCAG TTATACTCTA 840
AGGGAATATT CTCTGGCCCA TGTTCAACAT CATTGGACCA CGCAGTGCTC ATCGTAGGAT 900
ACGGTTCAAA GAACGGTGTT GATTACTGGA TCGTGAAGAA CTCTTGGGGA AAGAGTTGGG 960
GAATGGATGG GTTTATCCAC ATGCAGCGTA ACACCGGCAA CGCAGAAGGA GTATGCGGAA 1020
TCAACATGCT GGCTTCATAT CCCATCAAGA CACATCCAAA CCCTCCTCCA CCGTCCCCTC 1080
CCGGCCCCAC GAAATGCAAC CTTTTCACCT ATTGTTCAGC TGATGAGACT TGTTGCTGTG 1140
CGAGAAACTT GTTTGGTTTG TGTTTCTCGT GGAAATGCTG CGAGCTAGAG TCTGCTGTGT 1200
GTTGCAAGGA TGGTCGTCAT TGTTGTCCGC GTGATTACCC CGTCTGTGAT ACCACCAGAA 1260
GTCTTTGCCT TAAGAAAACT GGCAATTTCA CAGAGATCAA GCCCTTCTGG AAGAAGAATG 1320
CGTCCAATAA ACTTGGCAAG TTCGAGGAAT GGGTTATGTA AGAGGAAGTT TTCAAACTCT 1380
TTCACACGGT AAGCCTCTTT GGATTCGTTT ATCTATAAGC TGAGAGATGA TTACTTTATA 1440
GCTGTTGTTG TGATATGTAT TATTAGTCTC TTATTTGGAT GTATACAAAC TTTTGAATCA 1500
ATAAAAGGTT ACTTGCAGGA CACAATAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1560
AAAAAAAAAA AAAAAAA 1577
```

Fig.13. 1/2

```
                    10        20        30        40        50        60
                    |         |         |         |         |         |
AAAAGCTCTT CTTGTTTTCGG TTTCCCTGTA AGGAAAAGAA ATGTCACCAT CATCGTCTTC 60
TTCCTTTTGTC TCTATCACTT TCTTCTCCCT TCTTCTAGTT TCTTCTCTGA GCTTCCCATC 120
ATCATCTTCC GATGACATCT CCGAGCTGTT CGACGCTTGG TGCCAGAGAC ACGGCAAAAC 180
GTACGCTTCG GAAGAAGAGA GACAACACAG GATTGAAATC TTTAGAGACA ATCACGACTT 240
CGTCACACGA CACAACGGCA TCGCTAACTC TACTTACTCT CTCTCACTCA ATGCCTTCGC 300
GGATCTGACT CACCACGAGT TCAAGGCCTC TCGTCTTGGA CTCTCTGCTT CTTCAGCTCC 360
GTTGCTGGTG GCTAAGGGAG AAAGTGTTGA GAACGTTGGG GGCAAAGTTC CAGATTCTGT 420
TGATTGGAGG AAGAAAGGAG CTGTTACTAA TGTCAAAGAT CAAGGAAGCT GCGGAGCGTG 480
TTGGTCTTTC TCGGCGACTG GAGCAATGGA AGGAATCAAC CAGATTGTAA CAGGAGATCT 540
CATCAGCCTC TCTGAGCAGG AACTAATTGA TTGTGACAAG TCCTACAACG ATGGATGCAA 600
TGGTGGTCTC ATGGACTACG CTTTTCAATT TGTCATTAAA AACCATGGAA TCGACACAGA 660
GAAAGATTAT CCTTATCAAG AACGTGATGG CACCTGTAAA AAAGATAAGT TGAAAAGAAA 720
GGTTGTGACA ATTGATAGCT ATGCTGGCGT AAAATCAAAC GACGAGAAAG CGTTACTGGA 780
```

Fig.13. 2/2

```
AGCTGTAGCG GCTCAGCCAG TTAGTGTTGG CATCTGTGGC AGCGAGAGAG CGTTTCAGCT 840
ATACTCTAAG GGAATATTCT CTGGCCCATG TTCAACATCA TTGGACCACG CAGTgCTCAT 900
CGTAGGATAC GGTTCACAGA ACGGTGTTGA TTACTGGATC GTGAAGAACT CTTGGGGAAA 960
GAGTTGGGGT ATGGATGGGT TTATGCACAT GCAGCGTAAC ACCGGCAACT CGGAAGGAGT 1020
ATGTGGAATC AATATGCTCG CTTCGTATCC CATCAAGACA CATCCAAACC CTCCTCCACC 1080
GTCCCCTTCC GGCCCCACGA AATGCAACCT TTTCACCTAT TGTGCAGCTG ATGAGACTTG 1140
TTGCTGTGCG AGAAACTTGT TTGGTTTGTG TTTCTCGTGG AAATGCCGCG AGCTAGAGTC 1200
TGCTGTGTGT TGTAAGGATG GTCGTCATTG TTGTCCTCGT GATTACCCCG TCTGTGATAC 1260
AACCAGAAGT CTTTGCCTAA AGAAAACTGG CAATTTCACA GAGATCAAAC CCTTCTGGAA 1320
GAAGAATGCG TCCAATAAAC TTGGCAAGTT CGAGGAATGG GTTATGTAAG AGAAGTTTTT 1380
TAAACTCTTC CACACGGAAG CCTCTTTGGA TTCGTTATGT ATAAGCTGAG AGATGATTAT 1440
TTTATAGCTG TTGTTGTGAT ATGTATTATT AGTATCTCAT TTGGATGTAT ACAAACTTTT 1500
GAATCAATAA AGGGTATCTG CAGGACACAT TAAATAAAAA AAAAAAAAAA AAA 1553
```

Fig.14. 1/2

```
                    10        20        30        40        50        60
                    |         |         |         |         |         |
CAACTATCAA AACTAGAAAA CCAACAAAAC AAACATACAA TATGGCTTCG TCAGCGAAAC 60
TCCTCTCTTT ACTTCTCTTG TACGTCTTCA TTTCATTAGC CTCCAGTGAT GAGTCCATCA 120
TCAACGACAA CCATCTCATT CTTCCATCTG ACCGCTCGTG GAGAACCGAT GAAGAAGTGA 180
TGTCCATCTA CTTAAAATGG TCCTTGGAGC ACGGGAAAAG TAACAGCAAC AGCAACGGTA 240
TTATCAACCA ACAAGATGAA AGATTCAATA TTTTCAAAGA CAACCTAAGA TTCATCGATC 300
TACACAACGA GAACAACAAG AACGCTACTT ACAAGCTTGG TCTAACCATA TTCGCTGATC 360
TCACTAACGA TGAGTACCGG AGTTTATACC TCGGGGCAAG AACCGAGCCT GTCCGCCGCA 420
TCACTAAGGC CAAGAACGTT AACATGAAAT ACTCAGCCGC AGTAAACGAC GTGGAGGTTC 480
CGGAGACGGT TGACTGGAGA CAGAAAGGAG CCGTTAATGC CATTAAAAAC CAAGGATCTT 540
GCGGAAGTTG TTGGGCGTTT TCAACAGCTG CAGCAGTAGA AGGCATAAAC AAGATCGTAA 600
CAGGAGAGCT CATATCTCTG TCCGAACAAG AACTTGTCGA CTGCGACAAA TCATACAACC 660
```

Fig.14. 2/2

```
AAGGCTGTAA CGGCGGTCTA ATGGATTATG CTTTTCAATT CATCATGAAA AACGGCGGAT 720
TAAACACCGA GCAAGACTAT CCTTACCACG GAACCAATGG CAAATGCAAC TCTTTACTTA 780
AAAATTCGAG AGTTGTGACT ATCGATGGAT ACGAAGATGT TCCTAGTAAA GATGAAACCG 840
CGTTGAAGAG AGCAGTTTCG TACCAGCCTG TGAGTGTTGC TATTGATGCT GGTGGAAGAG 900
CTTTCCAACA TTACCAATCT GGAATCTTCA CTGGAAAGTG TGGTACGACT ATGGATCACG 960
CTGTTGTGGC GGTTGGTTAT GGATCAGAGA ACGGTGTTGA CTATTGGATT GTACGTAACT 1020
CTTGGGGTAC AAGCTGGGGA GAGGATGGTT ACATTAGGAT GGAGAGAAAC GTGGCGTCCA 1080
AATCCGGTAA GTGTGGGATT GCGATTGAAG CCTCGTATCC GGTTAAGTAC AGCCCAAACC 1140
CGGTTCGTGG AACCAGCAGT GTTTGAAGTT AACAAAAAGA ATCTCATGCA GTAATCAAAT 1200
TGGGATTGTT ATAAGTTAAA TTAATCTTGT ATTATTGTTT GTATGTATAG TATTTCGAAA 1260
AAAATTGATT CACCATAGGG ATTTAATCTG TATAAATCTC TATGTTGGTC AATATCATTT 1320
CATTCAAAGA ATATTTGCTT TGGCTTGATT ATGTATTAAG AGAAATATAA TAAAAAAAAA 1380
AAAAAAAAAA 1390
```

Fig.15. 1/2

```
         10        20        30        40        50        60
         |         |         |         |         |         |
CAACTATCAA AACTAGAAAA CCAACAAAAC AAACATACAA TATGGCTTCC TCACCAAAAC 60
TCCTCTCTTT ACTTCTCTTA TACGTCTTCG TTTCGTTAGC CTCCGGTTAT GAGTCCATCA 120
TCAGTGACAA CCATCTCAGT CTTCCATCTG ACCGTTCGTG GAGAACCGAT GAAGAAGTGA 180
TATCCATCTA CTTAAGATGG TCCTTGGAGC ACGGGAAAAG TAACAGCAAC AGCAACGGTA 240
TTATCAACCA ACAAGACGAA AGATTCAATA TTTTCAAAGA CAACCTAAGA TTCATCGATC 300
TACACAACGA GAACAACAAG AACGCTACTT ACAAGCTTGG TCTAACCATA TTCGCTGATC 360
TCACTAACGA TGAGTACCGG AGTTTATACC TCGGGGCAAG AACCGAGCCT GTCCGCCGCA 420
TCACTAAGGC CAAGAACGTT AACATGAAAT ACTCAGCCGC AGTAAACGAC GTGGAGGTTC 480
CGGAGACGGT TGACTGGAGA AAGAAAGGAG CCGTTAATGC CATTAAAGAC CAAGGAACTT 540
GCGGAAGTTG TTGGGCGTTT TCAACAGCTG CAGCAGTAGA AGGTATAAAC AAGATCGTAA 600
CAGGAGAACT CGTATCTTTG TCCGAACAAG AACTTGTCGA CTGCGACAAA TCGTACAACC 660
AAGGCTGTAA CGGCGGTCTA ATGGATTATG CTTTTCAATT CATAATGAAA AACGGCGGAT 720
```

Fig.15. 2/2

```
TAAACACCGA GAAAGACTAT CCTTACCACG GAACCAATGG CAAATGCAAC TCTTTACTTA 780
AGAATTCAAG AGTTGTAACT ATCGATGGAT ACGAAGATGT TCCTAGTAAA GATGAAACCG 840
CGTTGAAGAG AGCAGTTTCA TACCAGCCTG TGAGTGTTGC TATTGATGCT GGTGGAAGAG 900
CTTTCCAACA TTACCAATCT GGAATCTTCA CTGGAAAGTG TGGTACGAAT ATGGATCACG 960
CTGTGGTGGC GGTTGGTTAT GGGTCAGAGA ACGGCGTTGA CTATTGGATT GTACGTAACT 1020
CTTGGGGTAC ACGTTGGGGA GAGGATGGTT ACATTAGGAT GGAGAGAAAC GTGGCGTCTA 1080
AATCCGGTAA GTGTGGGATT GCGATAGAAG CCTCGTATCC GGTTAAGTAC AGCCCAAACC 1140
CGGTTCGTGG AACCAGCAGT GTTTGAAGTT AACAAAAAGA ATCTCATGCA GTAATCAAAT 1200
TGGGATTGTT ATAAGTTAAA TTAATCTTGT ATTATTGTTT GTATGTATAG TATTTCGGAA 1260
AAAAAAATGA TTCACCATAG GGATTTAATC TGTATAAATC TCTAGGTTGG TCAAAATATCA 1320
TTTCATTCAA AGAATATTTG McTTTGACTT GATTATGTAT TAAGAGAAAT ATAATAAAAT 1380
GGTATATTTC TCAACAGCAT TGGTTTCGCT GAAAAAAAAA AAAAAAAAAA AAAA 1434
```

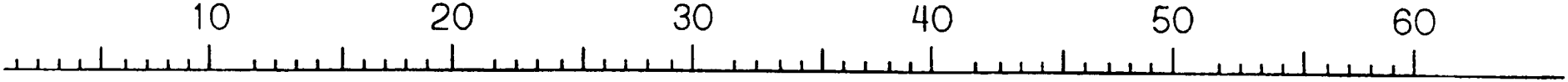
Fig.16. 1/2


```
AGAAAACCAA CAAAACAAAC ATACAATATG GCTTCGTCAG CGAAACTCCT CTCTTTACTT 60
CTCTTGTACG TCTTCATTTC ATTAGCCTCC AGTGATGAGT CCATCATCAA CGACAACCAT 120
CTCATTCTTC CATCTGACCG CTCGTGGAGA ACCGATGAAG AAGTGATGTC CATCTACTTA 180
AAATGGTCCT TGGAGCACGG GAAAAGTAAC AGCAACAGCA ACGGTATTAT CAACCAACAA 240
GATGAAAGAT TCAATATTTT CAAAGACAAC CTAAGATTCA TCGATCTACA CAACGAGAAC 300
AACAAGAACG CTACTTACAA GCTTGGTCTA ACCATATTCG CTGATCTCAC TAACGATGAG 360
TACCGGAGTT TATACCTCGG GGCAAGAACC GAGCCTGTCC GCCGCATCAC TAAGGCCAAG 420
AACGTTAACA TGAAATACTC AGCCGCAGTA AACGACGTGG AGGTTCCGGA GACGGTTGAC 480
TGGAGACAGA AAGGAGCCGT TAATGCCATT AAAAACCAAG GATCTTGCGG AAGTTGTTGG 540
GCGTTTTCAA CAGCTGCAGC AGTAGAAGGC ATAAACAAGA TCGTAACAGG AGAGCTCATA 600
TCTCTGTCCG AACAAGAACT TGTCGACTGC GACAAATCAT ACAACCAAGG CTGTAACGGC 660
GGTCTAATGG ATTATGCTTT TCAATTCATC ATGAAAAACG GCGGATTAAA CACCGAGCAA 720
```

Fig.16. 2/2

```
GACTATCCTT ACCACGGAAC CAATGGCAAA TGCAACTCTT TACTTAAAAA TTCGAGAGTT 780
GTGACTATCG ATGGATACGA AGATGTTCCT AGTAAAGATG AAACCGCGTT GAAGAGAGCA 840
GTTTCGTACC AGCCTGTGAG TGTTGCTATT GATGCTGGTG GAAGAGCTTT CCAACATTAC 900
CAATCTGGAA TCTTCACTGG AAAGTGTGGT ACGACTATGG ATCACGCTGT TGTGGCGGTT 960
GGTTATGGAT CAGAGAACGG TGTTGACTAT TGGATTGTAC GTAACTCTTG GGGTACAAGC 1020
TGGGGAGAGG ATGGTTACAT TAGGATGGAG AGAAACGTGG CGTCCAAATC CGGTAAGTGT 1080
GGGATTGCGA TTGAAGCCTC GTATCCGGTT AAGTACAGCC CAAACCCGGT TCGTGGAACC 1140
AGCAGTGTTT GAAGTTAACA AAAAGAATCT CATGCAGTAA TCAAATTGGG ATTGTTATAA 1200
GTTAAATTAA TCTTGTATTA TTGTTTGTAT GTATAGTATT TCGAAAAAAA TTGATTCACC 1260
ATAGGGATTT AATCTGTATA AATCTCTATG TTGGTCAATA TCATTTCATT CAAAGAATAT 1320
TTGCTTTGGC TTGATTATGT ATTAAGAGAA ATATAATAAA AATGATATAT TTCTCAAAAA 1380
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1440
A 1441
```

Fig.17. 1/2




```
CAGAACTAGA ACAACCAAGC CAAACATACA ATATGGCTTC CTCAACAAAA CTCATTTCTT 60
TACTTCTCCT ATACGTCGTC GTTTCATTAG CCTCCGGTGA TGAGTCCACT ACCATTAACA 120
ACCATCTCAA TCTTCCATCG GACGGCTCAT GGAGAACCGA TGAAGAAGTG AGGTCCATCT 180
ACTTACAGTG GTGTGCGGAG CACGGGAAAA CTAGCAACAA CAACGGTATC GTCAACCAAC 240
AAGACGAAAA GTTCAATATT TTCAAAGACA ACCTAAGGTT CATTGATCTA CACAATGAGA 300
ACAACAAGAA CGCTACTTAC AAGCTTGGTC TCACCATATT CTCTGATCTC ACTAACGATG 360
AGTACCGGAG GTTATACCTC GGGGCAAGAA CCGAGTCTGT CCGCCGCATC ACTAAGGCCA 420
AGAACGTTAA CATGAAATAC TCGGCCGCAG TAAACGACGT GGAGGTTCCG GAGACGGTTG 480
ATTGGAGACG GAAAGGAGCC GTTAATGCCA TTAAAAACCA AGGAACTTGC GGAAGTTGTT 540
GGGCGTTTTC GACAGCTGCA GCAGTAGAAG GTATAAACAA GATCGTAACA GGAGAACTCA 600
TATCTCTGTC CGAACAAGAA CTTGTCGACT GCGACAGATC CTACAACCAA GGCTGCAACG 660
GTGGTTTAAT GGACTATGCT TTTCAATTCA TCATGAAAAA CGGCGGTTTG AACACCGAGC 720
```

Fig.17. 2/2

```
AAGATTATCC TTACCGTGGT TCCAATGGAA AATGCAATTC TTTACTGAAG AATTCAAGAG 780
TTGTAACTAT TGATGGTTAC GAAGATGTTC CTACTGAAGA TGAAACGGCG TTGAAGAGAG 840
CAGTTTCATA CCAGCCCGTG AGTGTTGCCA TTGAAGCTGG TGGAAGAGTT TTCCAACATT 900
ACCAATCGGG GATCTTCACT GGAAAGTGTG GGACAAATCT AGATCATGCA GTGGTGGCTG 960
TTGGTTATGG TTCAGAGAAC GGTATTGACT ATTGGATTGT AAGGAACTCG TGGGGTACAC 1020
GTTGGGGAGA GGATGGTTAC ATTAGGATGG AGAGAAACTT GGCAAGGTCC AAGTCCGGCA 1080
AGTGTGGAAT TGCGGTTGAA GCCTCGTACC CGGTTAAGTA CAGTCCAAAC CCGGTTCGTG 1140
GAACCAGCAG TGTTTGAAGT TTTTAAAATA AAACTCAATT GGGAGTTTTA TAACTAAGAT 1200
TTAATCTCAT ATTATTGTTT GTATGTATAG TATATCAAAA AAGAAGGTAT TTGATCCAGC 1260
ATACGGATTT AGTCTGTATA AATCCTTATG TCGATCAATA TCATTTCGTT CAAAGAAAGA 1320
TTGATTTGGT TGTTTATGTA TTAAGAGAAG TATAATAAAA TGATATATTT CTCTTAAAAA 1380
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1440
AAAAAAAAAA AAAAAAAAAA AAAA 1474
```

Fig.18. 1/6

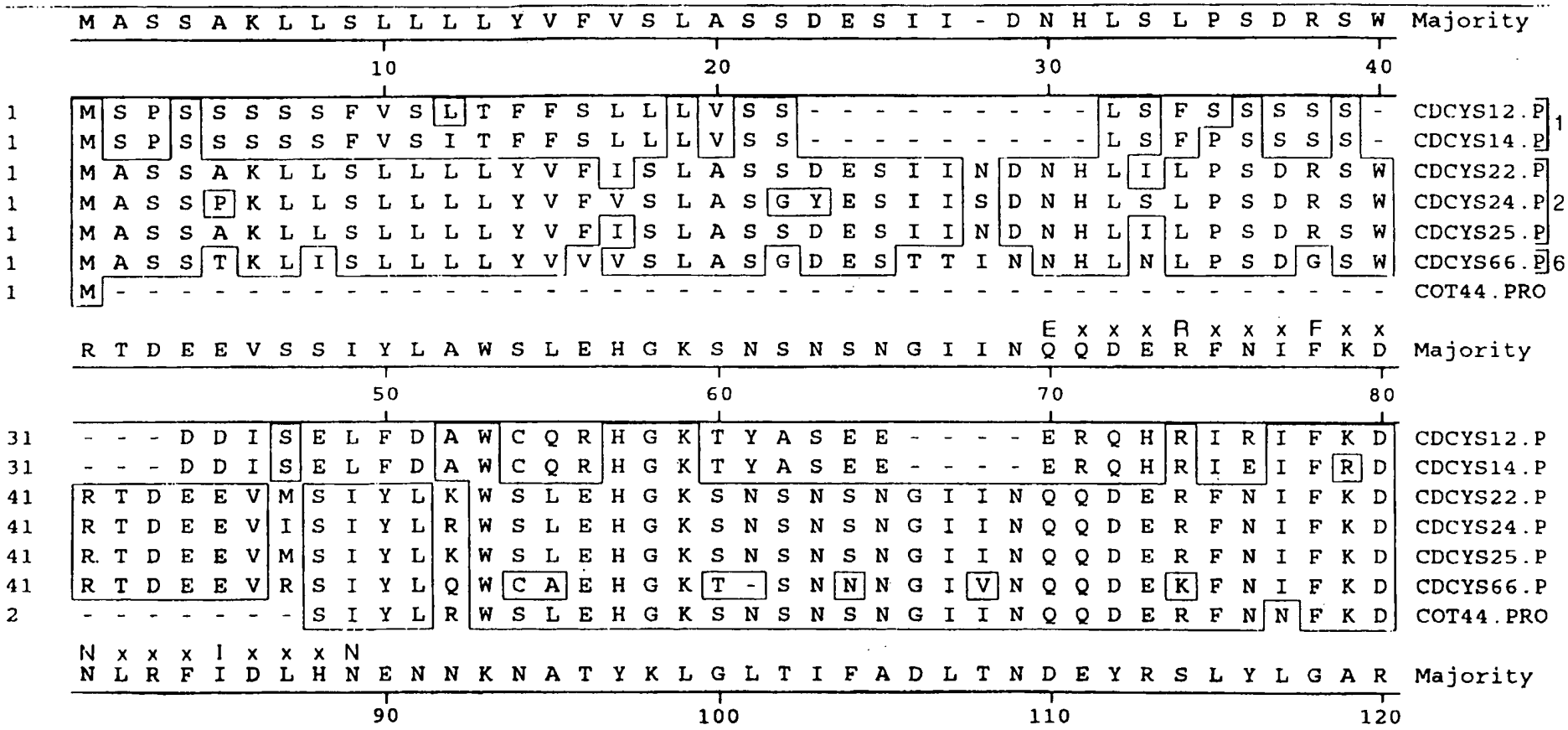

M A S S A K L L S L L L L Y V F V S L A S S D E S I I - D N H L S L P S D R S W Majority
10
20
30
40
1 M S P S S S S S F V S L T F F S L L L V S S - - - - - - - - - - L S F S S S S S - CDCYS12.P 1
1 M S P S S S S S F V S I T F F S L L L V S S - - - - - - - - - - L S F P S S S S - CDCYS14.P
1 M A S S A K L L S L L L L Y V F I S L A S S D E S I I N D N H L I L P S D R S W CDCYS22.P
1 M A S S P K L L S L L L L Y V F V S L A S G Y E S I I S D N H L S L P S D R S W CDCYS24.P 2
1 M A S S A K L L S L L L L Y V F I S L A S S D E S I I N D N H L I L P S D R S W CDCYS25.P
1 M A S S T K L I S L L L L Y V V V S L A S G D E S T T I N N H L N L P S D G S W CDCYS66.P 6
1 M - COT44.PRO
R T D E E V S S I Y L A W S L E H G K S N S N S N G I I N E x x x R x x x F x x Q Q D E R F N I F K D Majority
50
60
70
80
31 - - - D D I S E L F D A W C Q R H G K T Y A S E E - - - - E R Q H R I R I F K D CDCYS12.P
31 - - - D D I S E L F D A W C Q R H G K T Y A S E E - - - - E R Q H R I E I F R D CDCYS14.P
41 R T D E E V M S I Y L K W S L E H G K S N S N S N G I I N Q Q D E R F N I F K D CDCYS22.P
41 R T D E E V I S I Y L R W S L E H G K S N S N S N G I I N Q Q D E R F N I F K D CDCYS24.P
41 R T D E E V M S I Y L K W S L E H G K S N S N S N G I I N Q Q D E R F N I F K D CDCYS25.P
41 R T D E E V R S I Y L Q W C A E H G K T - S N N N G I V N Q Q D E K F N I F K D CDCYS66.P
2 - - - - - - - S I Y L R W S L E H G K S N S N S N G I I N Q Q D E R F N N F K D COT44.PRO
N x x x I x x x N
N L R F I D L H N E N N K N A T Y K L G L T I F A D L T N D E Y R S L Y L G A R Majority
90
100
110
120

Fig.18. 2/6

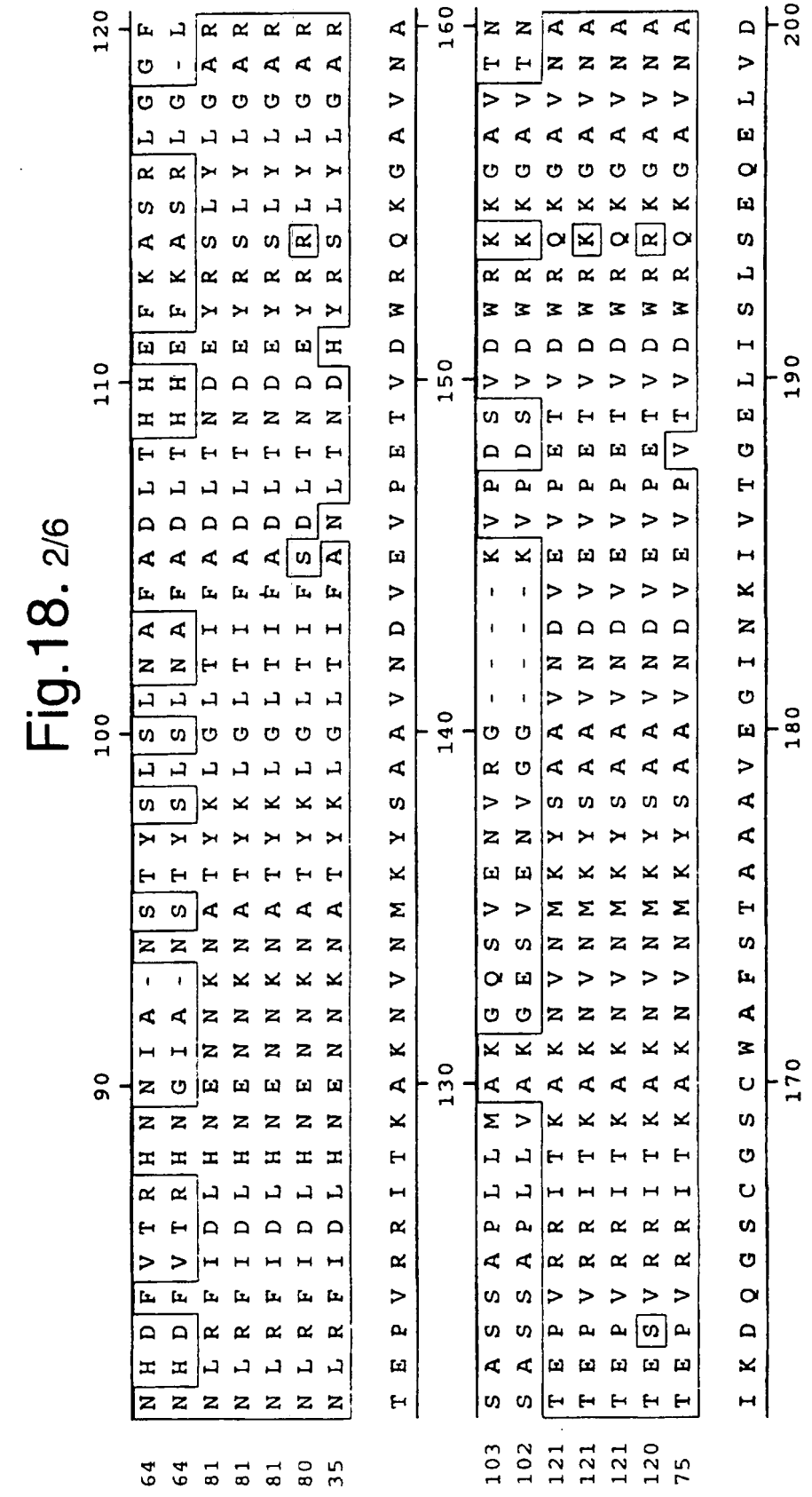

90
100
110
120
130
140
150
160
170
180
190
200
CDCYS12.P
CDCYS14.P
CDCYS22.P
CDCYS24.P
CDCYS25.P
CDCYS66.P
COT44.PRO
Majority

Fig.18. 3/6

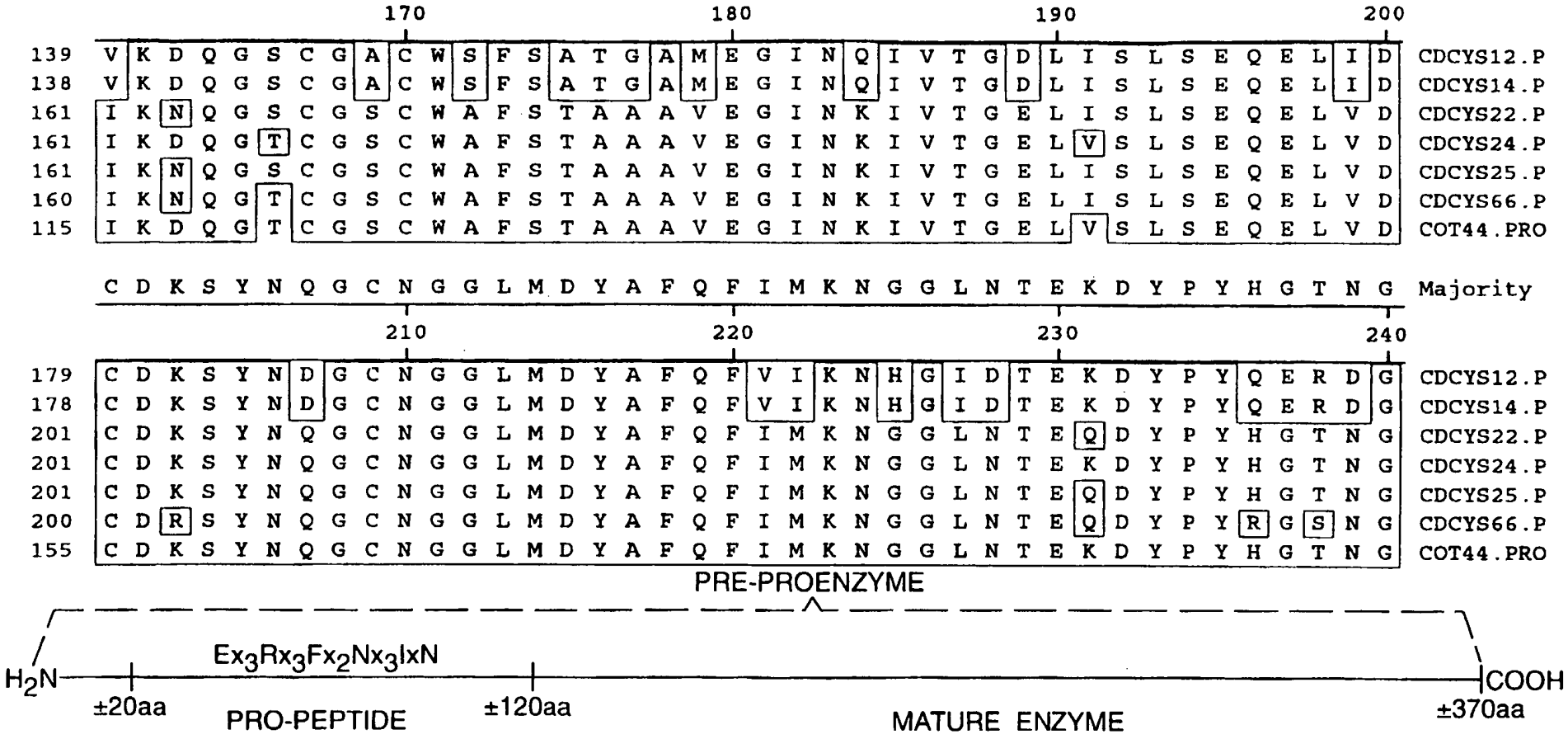

170 180 190 200
139 V K D Q G S C G A C W S F S A T G A M E G I N Q I V T G D L I S L S E Q E L I D CDCYS12.P
138 V K D Q G S C G A C W S F S A T G A M E G I N Q I V T G D L I S L S E Q E L I D CDCYS14.P
161 I K N Q G S C G S C W A F S T A A A V E G I N K I V T G E L I S L S E Q E L V D CDCYS22.P
161 I K D Q G T C G S C W A F S T A A A V E G I N K I V T G E L V S L S E Q E L V D CDCYS24.P
161 I K N Q G S C G S C W A F S T A A A V E G I N K I V T G E L I S L S E Q E L V D CDCYS25.P
160 I K N Q G T C G S C W A F S T A A A V E G I N K I V T G E L I S L S E Q E L V D CDCYS66.P
115 I K D Q G T C G S C W A F S T A A A V E G I N K I V T G E L V S L S E Q E L V D COT44.PRO
C D K S Y N Q G C N G G L M D Y A F Q F I M K N G G L N T E K D Y P Y H G T N G Majority
210 220 230 240
179 C D K S Y N D G C N G G L M D Y A F Q F V I K N H G I D T E K D Y P Y Q E R D G CDCYS12.P
178 C D K S Y N D G C N G G L M D Y A F Q F V I K N H G I D T E K D Y P Y Q E R D G CDCYS14.P
201 C D K S Y N Q G C N G G L M D Y A F Q F I M K N G G L N T E Q D Y P Y H G T N G CDCYS22.P
201 C D K S Y N Q G C N G G L M D Y A F Q F I M K N G G L N T E K D Y P Y H G T N G CDCYS24.P
201 C D K S Y N Q G C N G G L M D Y A F Q F I M K N G G L N T E Q D Y P Y H G T N G CDCYS25.P
200 C D R S Y N Q G C N G G L M D Y A F Q F I M K N G G L N T E Q D Y P Y R G S N G CDCYS66.P
155 C D K S Y N Q G C N G G L M D Y A F Q F I M K N G G L N T E K D Y P Y H G T N G COT44.PRO
PRE-PROENZYME
H2N
Ex3Rx3Fx2Nx3IxN
±20aa
PRO-PEPTIDE
±120aa
MATURE ENZYME
COOH
±370aa

Fig.18. 4/6

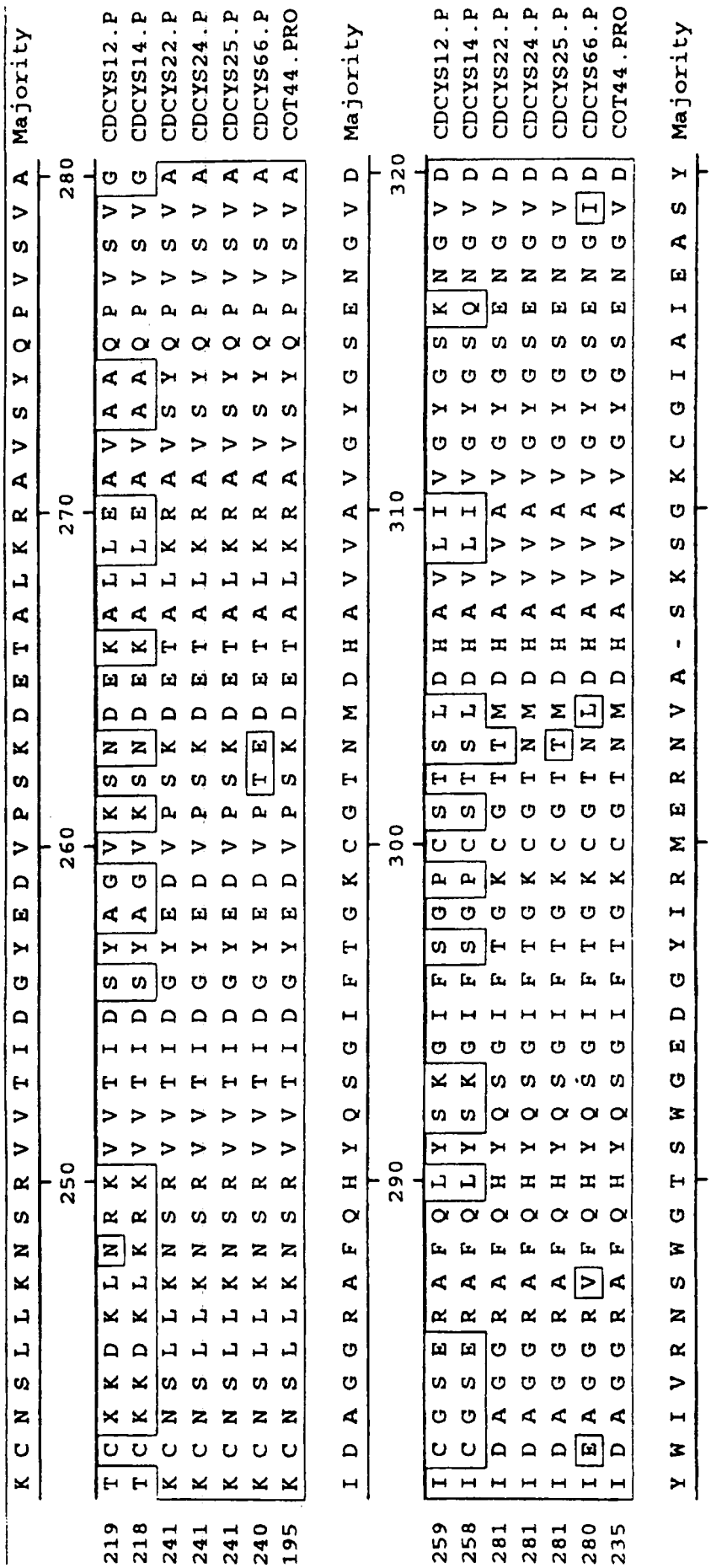

Majority
CDCYS12.P
CDCYS14.P
CDCYS22.P
CDCYS24.P
CDCYS25.P
CDCYS66.P
COT44.PRO
250
260
270
280
290
300
310
320
330
340
350
360

Fig.18. 5/6

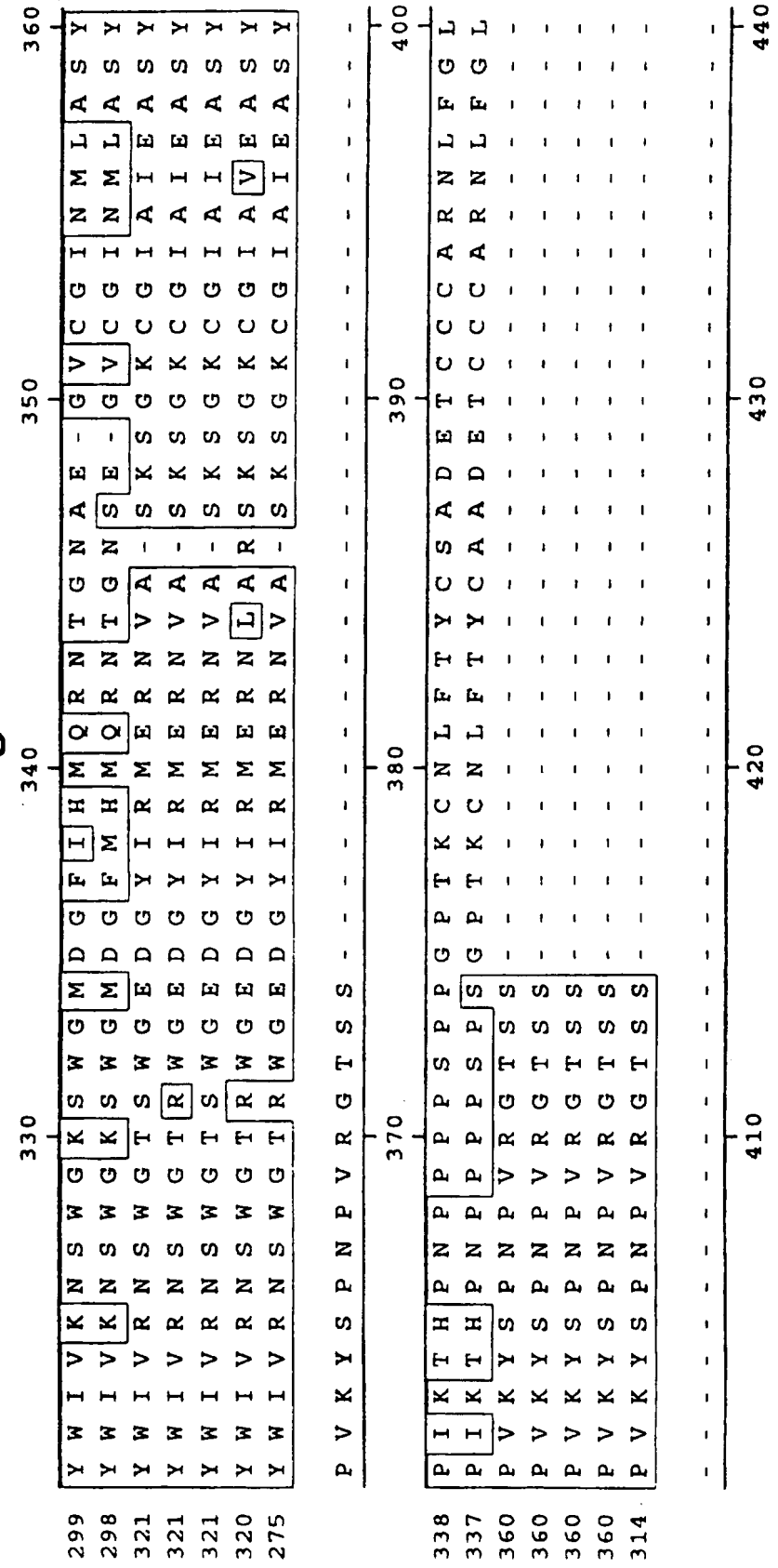
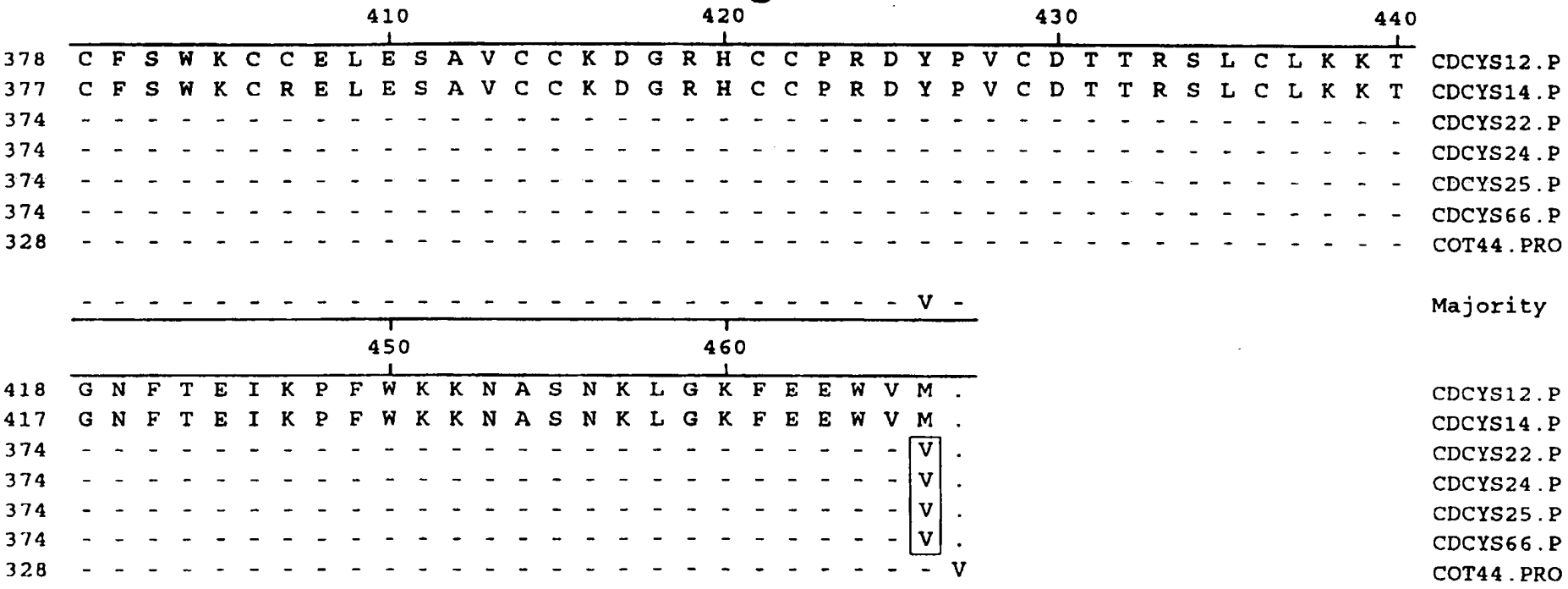

330
340
350
360
370
380
390
400
410
420
430
440
CDCYS12.P
CDCYS14.P
CDCYS22.P
CDCYS24.P
CDCYS25.P
CDCYS66.P
COT44.PRO
Majority

Fig.18. 6/6



Decoration 'Decoration #1': Box residues that match the Consensus exactly.

Fig.19. 1/2

Contig 2:

Contig Length:	1644 bases
Average Length/Sequence:	449 bases
Total Sequence Length:	9448 bases
Top Strand:	11 sequences
Bottom Strand:	10 sequences
Total:	21 sequences

BamII 10 20 30 40 50 60

```
GATCTACTTC GGCTAAAATT ACAGTCTCAA CAACTACACA AGTAGCTCAG AGAATCAGCA 60
GAAACAGTTT GCTACCCAGA AGAACCCTAA TTCAACAGGG GAGATAAAAA AAAAGTAAAA 120
AGAGTACACT GAGAGAATAA GGATGATCAC CTCGAGTTTT CTGATGAAGA AAGGGAACCA 180
AAGTGAAGCA TTTGATGATT GTCTTCTTCG ATTGATGAGC TTCTTCACAT CTAAAAAAGT 240
ATACTTTTTT CTCTCTCGAT GATTGATGAT CAGATAGAAG AAGAAAAAGA TAAACCTTAG 300
ATTTTTTTTT GTTGTTCACT CTTCACTTGG TCTTCTACTT CTTCTTCTTC CACCTTTGTT 360
TGTCTCTACC GTTTTGAATC AAGCGAGATT ATGAAAGGAC AACTCATCAT TATCACCATT 420
GATATCTTTA ATCCTTTATT TATATATTTA TTTACCCATT TAATAGTTTT TTATGCTTAG 480
```

Fig.19. 2/2

```
TTATGTATTT AGAGAAATTA CTTTATACTG TTTAGCCAGG AATACATATA TCAGTTAAAC 540
AATAGTGACC TGTTAATTAC TAAAATTTAA TAAAGTAGAG ATGTCACCGA ATATTGTGAC 600
ATTAATAAGA AGCAGTTTTC AAACCTTTTT AGCCTACCTA ATATAACCTG ATATTCAAAC 660
TTTTGATCTA CGACCTTATT TCAACTCTAG TAGTTGTAGA TACTTACAAA ATAAAATGTC 720
ATCGATTTCA AGTAAACTAA ACATGCATTT ACATGGGACA TATTCTCTAA TGTTATTATA 780
CCTTCTGATA AAACAACAAT AAATGTTCTT AGAATTGGAA AAATACTATT TTTTCTAAAG 840
AGAAAAAGGC ACGTTCGTAT TTGCTGATTA TTATATACAG TAGTAGTAAA AAGTAGCCTT 900
TACTTGTCGA CAGTTAGGTA AAGACTGAAC GCCACGCCAC ACTATCTTCT TCTTTTCAAC 960
AGTGTGAAGA TGTGTTTGTT TCTTTGTCAC TCGTTCTATT CTATCATCAT CATCTTAGGT 1020
CACTAGCCAC ACTTATGTTT TTTCTAAAGT ATACACTGGA AAGATTGGTA AATGTATTTG 1080
ATAATATATA TGTGATTAAT GATGTAGTAA TTCTAAAACT AAAGCATATT TCTTGGTTAC 1140
TTACTAGGTA CTACTACTCT CTGGTCTCCG CGTGATCTTA TATTTATTAT ACTAATTGAA 1200
ATTAAAAAGC ATATACTAAA AAGGGTTAAT GCCATAAAGT CGTAAGTAGG TCCAACAAGG 1260
AGTGGTCTTA TTAACCTAAA AAAGGTTGAA AGTAGCTTTC TTTTTGCTTA CAAAAGTATA 1320
TGCTATGTTT ACATTTAACA ATGAATTTAT TGTTGACCAA AAAAACAAAC AATGATTTAT 1380
TTAGCTAAGA GTAGCATTTG AATTTTGATT GTAATGGGCC TAGCCTGTCT GGCCCAATCA 1440
TTCATAACCG GGTCTAATCA TTAATAAAAA GCCCTTCAGC ATTCAACGCT AAACGCTTAT 1500
TACAAACGCT AGCCGCGCGT TTTGTAAACG TTGTTGCATC CTTACTGGCT AAAGTCTCCT 1560
TATTACTAGG ACAAACTATC TAATCCACGA CGACGACCAA ACAAAAGCTC TTCTTGTTTT 1620
GGTTTCCCTG TAAGAAAAGA AATG 1644
                    START
```

Fig.20A. 1/3

Contig 5:

Contig Length:	3997 bases
Average Length/Sequence:	452 bases
Total Sequence Length:	21697 bases
Top Strand:	28 sequences
Bottom Strand:	20 sequences
Total:	48 sequences

HindIII 10 20 30 40 50 60

```
AAGCTTGATG GGGTTGTGAT GTTTGCGGAT GATAGTAACA TGCATAGTAT GGAGTTTTTC 60
GATGAGATTC AGAACGTGAA GTGGTTCGGT GCTGTTTCCG TTGGGATATT AGCGCATTCG 120
GGGAATGCGG AAGAGATGGT TATGTCGATG GATAAGAGAA GAGAGATGGA GGAAGAAGAG 180
AGCTCTTCGT TATCTTCGTT ACCAGTACAA GGTCCTGCGT GTAACGCGAC CGATAAGCTG 240
```

Fig.20A. 2/3

```
ATCGGTTGGC ATGTTTTCAA TACGTTGCCA TACGCGGGGA AGAGTGCGGT TTACATAGAC 300
GATGTAGCTG CGGTTTTGCC GCAGAAGCTG GAGTGGTGTG GGTTTGTATT GAACTCGAGG 360
ATTCTTTGGG ATGAGGCTGA GAGTAAGCCG GAGTGGGTTA AGGAGTTTGG GTTGTTGAAC 420
GAGAACGAAG gCGTGGAGAG TCCTTTGTCT CTGTTGAATG ATCCTTCGAT GGTTGAGCCT 480
CTTGGAAGCT GTGGAAGACA GGTTCTGCTT TGGTGGCTTC GTGTCGAAGC ACGCGCTGAT 540
AGCAAGTTCC CTCCCGGGTA TATGCCTTCT TTTGCTCTCA AGATGTTAAC TAATTAGCTC 600
AATGTTCAAA CTAGGAGTTA GTTAGTGATG TTTGGTTCTG TTTTTGTTAT ATGAGTGCAG 660
ATGGGTGATT GATCCTCCGT TAGAAATCAC AGTGGCGGCT AAACGAACGC CATGGCCAGA 720
TGTTCCACCT GAGCCACCTA CTAAAAAGAA AGATCAAATG TCATTATCCC AAGGCAACAA 780
CAACGTGGTG GTGATACCGA AGCACCAGCA GCAGCAGCAG CAACAGCAAC GTTCTAGCAA 840
AGTGCGGAAA CCGAAACGCA GAAGTAAGAG AAATAAACAC GAAGCTAAAC CAACTGATAC 900
GACAACACAA GTTTCTTCTT CCACTAAACA TCATCAAGAA AGAAACTGAG GAAGAAGAAA 960
GATCATTTCT ACTATTTTAT TATCATTCAT TTGTTTGCCA AAGTTTTATA GAGAAATGTC 1020
AAGAGATCAT CTTATTCTCC TCCAGATACC GCGAATAGTA AGAATCAACG GTGAGGAAGC 1080
AAGAACAAAG GCTTAGATTT TATGATGATG GAGCCCCCAC AGCAAAAGGG TATACGGGTT 1140
TTTAGGGGAA TTGATGATCA TTCGTTTATT TTCGATATAT CTTTCTTTTG GTCTTTATAA 1200
```

Fig.20A. 3/3

```
GAACTTATCT GTATTAGTAA CAAAGAATTA TTTGTTTCAT TGTTCTTGAA GTTGCAAAAA 1260
AAGGCTTAGA AACAAAAACT TTGGGTTGTA TAGTTTCTTT TGTAAGCATA TTTCTCAATC 1320
CATTTCTGTT TTATCCCCCA ATTAATCCAC ACTATGGACC TAAAAGTTCA AAGCCTCAAT 1380
TTACTCCTGC GAGTTCGTTT CACAGAAATT AATTACACTT TTTTATTATT CTTCATAAGG 1440
AAACAATACA TATGTATTTA TATAGTATTT TCTAAAGAAA TATTGGAGAT GAAGACATTG 1500
GAAAAGATAC CTTCGTAAAA AAGAAATTGTG AAGAAAGTAA ATACAAAAACA GAATCATTCT 1560
TTTGTGAAAG TTGAGACAAA GACTGAAAAC AGATTCATAA CTGAAATAAT GGAGGATTGG 1620
GCGGTTTTGT CAGAACATCA CATGTCTTCT TTAACCACTT TTCCCTCACT CTATTAGATT 1680
TTTTTCTTAG TACACTGCGG TCATAGCCTA TAGGTATGTA CTTTGACCAC TATAAACAGG 1740
ATTGAGTTAG GGTTTCTTCT AATTATGGGA TTTTATTACG TTTTCAACCT CTATTTTCTT 1800
CGGATTAGGT TTATTTAGAT GTTTTTTGCT TCATTTCTTT AACTATTTAG TTTTAGAGAG 1860
GAAACACAAG ATGTTAAACC CATTAGTGAA CCAAAAATTTA AGCAGGTGAC TAGAGTAAAC 1920
CAAACATATC TAATAGTATG GAATTTTGTC TTTACTACTT AATATGAGGT AATCAATCCG 1980
GATAGAAGTG AAATTAACGA AAAATATCCT ATTTATATGT TTACTTATAA AATCAAATTT 2040
ATATTATGCG ATTTTTAACA TTCTTTGGAA CAGTATTAGT TCAGTTGCTT AAACGAAAAC 2100
```

Fig.20B. 1/2

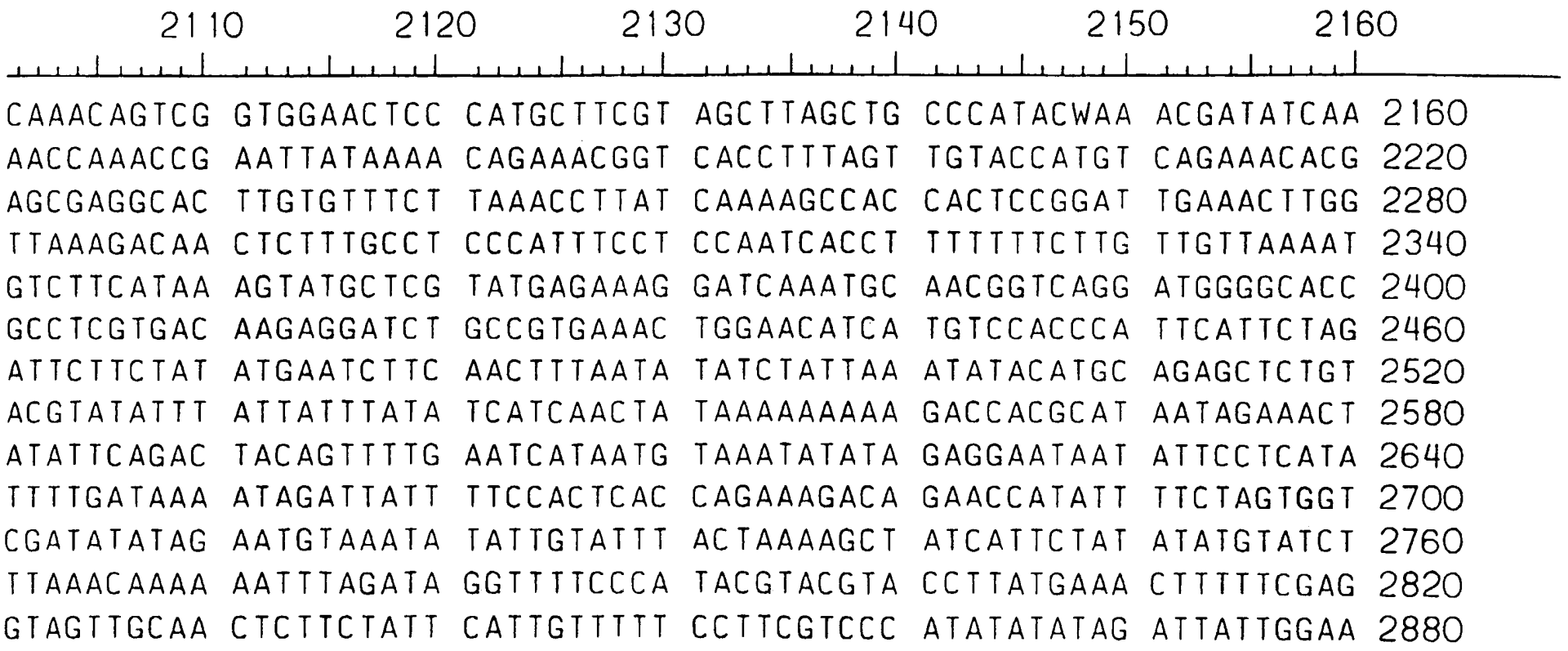

2110
2120
2130
2140
2150
2160
CAAACAGTCG GTGGAACTCC CATGCTTCGT AGCTTAGCTG CCCATACWAA ACGATATCAA 2160
AACCAAACCG AATTATAAAA CAGAAACGGT CACCTTTAGT TGTACCATGT CAGAAACACG 2220
AGCGAGGCAC TTGTGTTTCT TAAACCTTAT CAAAAGCCAC CACTCCGGAT TGAAACTTGG 2280
TTAAAGACAA CTCTTTGCCT CCCATTTCCT CCAATCACCT TTTTTTCTTG TTGTTAAAAT 2340
GTCTTCATAA AGTATGCTCG TATGAGAAAG GATCAAATGC AACGGTCAGG ATGGGGCACC 2400
GCCTCGTGAC AAGAGGATCT GCCGTGAAAC TGGAACATCA TGTCCACCCA TTCATTCTAG 2460
ATTCTTCTAT ATGAATCTTC AACTTTAATA TATCTATTAA ATATACATGC AGAGCTCTGT 2520
ACGTATATTT ATTATTTATA TCATCAACTA TAAAAAAAAA GACCACGCAT AATAGAAACT 2580
ATATTCAGAC TACAGTTTTG AATCATAATG TAAATATATA GAGGAATAAT ATTCCTCATA 2640
TTTTGATAAA ATAGATTATT TTCCACTCAC CAGAAAGACA GAACCATATT TTCTAGTGGT 2700
CGATATATAG AATGTAAATA TATTGTATTT ACTAAAAGCT ATCATTCTAT ATATGTATCT 2760
TTAAACAAAA AATTTAGATA GGTTTTCCCA TACGTACGTA CCTTATGAAA CTTTTTCGAG 2820
GTAGTTGCAA CTCTTCTATT CATTGTTTTT CCTTCGTCCC ATATATATAG ATTATTGGAA 2880

Fig.20B. 2/2

```
ATTGATGTAA CGTAGATTCA GAAATTCACA TCACAGACAT CTATACCTAT TTATACGGCT 2940
CTCAAACCAT AATTGTCACA ATGCATGTCT GTGGATTTAA CTTTACAACT TACTAAATAC 3000
TCCAATTAGA TTCGACGTAA AGAAATTAAT CACACGAAAA AGATGCAGCT ATTTCGTTAA 3060
AAGTTGATCC TCAAGTGACA TCATGCTGCT TCGAAATGCT GAAAAGATTT AATGATCTTC 3120
ATGTTCCCAT GTGTTCGGAA TTGAACGACT CTGATGATAT GTGGTCGGTA GCCAAAAAAA 3180
ACGCGGTTCA AAGCGAATTG TTTGTTTTGC TATATACTAT TGACGTTATA AAGTCAAAAA 3240
TAAACCGCTT ACAAGATGGT TCTGACCGCT TCAAAAATAA ACGCTCTGTG TATATATATT 3300
ATCGTCTACT GTATGCAGGA TATATGGTAG CTATTTACGT CAAATATGAG TAGATCAGCG 3360
TTGATCAAAC ATAAACGTAA ATCTCTCGTG AGTCGATAGA TGTTGATGTA AACAACAAAC 3420
CACGGAAAAT AAATAACTGC AAAAAATATA AAAAGAATGG TGTTATTATT TAACCCAAAA 3480
ATTGTATTAT CATACTTACA CGAAGAAAAT TACGTACATC TACGAGATAA TTCCATGAGA 3540
AAAGAATAGT AAACTCTTTG AATTACAGAC GAATGGTTTG AACCCCGTAG ATCTTTGAGT 3600
TTGAATTGTA TTTAAGAGCA TGATACTACG ATGTTGTATT CTTTTTTAGA TCAGTGAATC 3660
GTAGTTTTTT CCTATTTACC ATCCGATTGT CCGAATAAAA TAATAATTAT CTCCAAAATT 3720
AATAAATTTA GATTTCTAAA ATCAAACATA AATATCTAGT AATTAAAAGT GAAGATATTA 3780
TATATTTAGT AAACAGTTGT CCATATTTCG GCAAAAAAAA TCCATTTGAA ACGTCCACTA 3840
TCTTGCATGC CCATTATTAC TTTTTATTAT AATACCAACT TGAAATATTA AAATACCTAA 3900
ACTTTGGTTA TAAATAGTTT CACATTCTTG TCCCACCAAA AATCAAGCCA TACCAACTAT 3960
CAAAACTAGA AAACCAACAA AACAAACATA CAATATG 3997
                                    START
```

Fig.21. 1/3

Contig 2:

Contig Length:	2000 bases
Average Length/Sequence:	413 bases
Total Sequence Length:	9920 bases
Top Strand:	14 sequences
Bottom Strand:	10 sequences
Total:	24 sequences

BamHI 10 20 30 40 50 60

```
GGATCCCACA CATAACTGTA ATGTTTCAAT ACTCACGTGT AACTTTGATC ATCGAAACCT 60
ATTTAGTAAA ATCCGCATTC TGGCCCAATA AAACTTATAA GTGGGCTAAA TCTCTTTTGT 120
ATGTATCTAG GATTTTTATA GTACTATGTC TCCACCGATA AACCGAAGCG TTACCCTTTT 180
AGCTTATCAA AAAAAATTCT GAAACTTTTC ATTTTCACCT CTCTATATCT CCAACGATCA 240
```

Fig.21. 2/3

```
GTTATGGTAC CGTTTCACCT CTGAAACGAT CCGTCTCAGT ATATATAATT CCTCAAACAA 300
ACCTTGAAAC CCATATCTCT TATTAAATTA CTCCTAAATT GAATTGTCGC GGCTTTTAGC 360
CAACCTTCGA AGATGTCAGC CTCCAGTGTT GTTGTTGCTT AATCGGTCGT TGTCCCCGCC 420
ACCTTTCTTC GCCTAGAACA TAGTACCCAG GAAAAGGTAT TGACAATTTG ACTTTCTATG 480
GAATACTATG TACCATATTT GTTTAATCAA TATTCAAGCT ATCAGTTTTA ATACTTCAAG 540
CGTATGTTTC TCTGGGCAGA ATTCTGAGAT CCACGGTTCT ATCCCAGTTC TTGTGAATCA 600
CTACCATCCA TCGCTACGAT AGGGGGTCGA TTGTGAAAGT TGATTGATCG CTTTGAAGTT 660
GTCCGATTAA CTAATATGAT GTAAACTCTG TTCTGATGAG CTATAACGGA AAGTTATTTG 720
GTTGCTATGA CTAACAGAGG CTGCAAGGCT TGGTACGTCT CAGGCGGAGC TGAAGGACGG 780
TACAACTAGA GCGCCTATTG TCTGGTTTTT GTCGTGCCAA CGTGCTATGA AACTTAAGCC 840
TTCCTTGGAG AATCCAGATC ACTTCGAAAC GTCGGCTGTC ATTTTCAACA GGTACCACCA 900
TGATGTAAAT TTTTTTATGT TTGTGGTGTG GTTCGGCTCA AGATTGAAAC TTTTGCTGAT 960
GAAATAATGA TTTTGTGGAT GCATGACAAT TAGATTTACG AGTCTACAAA CTGTTCATTG 1020
TACATTGGCA GAGAAAAAAT GGCTATGTAT GGTTCAGTTG TAACACCGGT GTGGTGCAAT 1080
GAGGATGGAC ATGGTAACCA AAGGTGTGCA TAATGTTATT GATTTCCTCA GCGATGATTC 1140
CCCTAACATG GATGTGATCG GAATCTCTGG TGAGTTTCTT TTGTATGTTA TTGGTTTGAT 1200
```

Fig.21. 3/3

```
TTTAAAACAA TGTCTTTTAG CATAGTATCA TCTTGAGACT TAAAAACTTG AAATTTATTG 1260
GTCAGATAAC TTATTCTCCG ACATTAATAA ACCTGCTGCG GTGAACTGCA TCGAGGGACA 1320
TGGCAGTAAT TAGAAGAGAG ATAGGGAATA AGGTGTTGAA AACGAGTGTG TCTGGCTTGG 1380
TGGAGCTCAA TATACTCAAG AACCTCACCG TCTATGTTGT TGCAGGCTCT CTAGGTCGAT 1440
TCAAATCTCA TGCCAGCAAC ATAGTGTTTG ATGTATTCAT AGCTACTTGC CAAGATCCAG 1500
CCCAAAACAT GGAGAGCTCT CAAATTGATG GAAGCCATTA ATAGCAATCA CCGTACCTTC 1560
ACTATCTCCA CCATCACTGA AGATACCATC ATAGAGAACT ATGGAAAAAT CGTCTAATTA 1620
TACTTGGGTA TATATTATAG ATCAGCAAAC CGTATTTATT GTTATCTTGA TAGTTGATAT 1680
AGTATATAAG TAACTAAATT TTCTGAAATT ATTAGAAAAT ACATAAATAT CTCCCTGCCT 1740
ATTATCACAC AACGTTCTTA CGTGGGGAAT GAAGATATTT AAGGTGTAAA ATTAATTTCA 1800
TTCATATTTC CGGCAATATC CATTTGAACC GTCCACCATC TTGCATGCCC ATTACTGCAT 1860
TTTATTATAA TAGAAAAGTA TACCAACTTG AAATATTAAA ACTCCAAGAA TTTGGTTATA 1920
AATAGCTCCT CTCAGCCTCC AACGAAATCA AGCCATATCA ACTATCAGAA CTAGAACAAC 1980
CAAGCCAAAC ATACAATATG 2000
                     START
```

Fig.22.

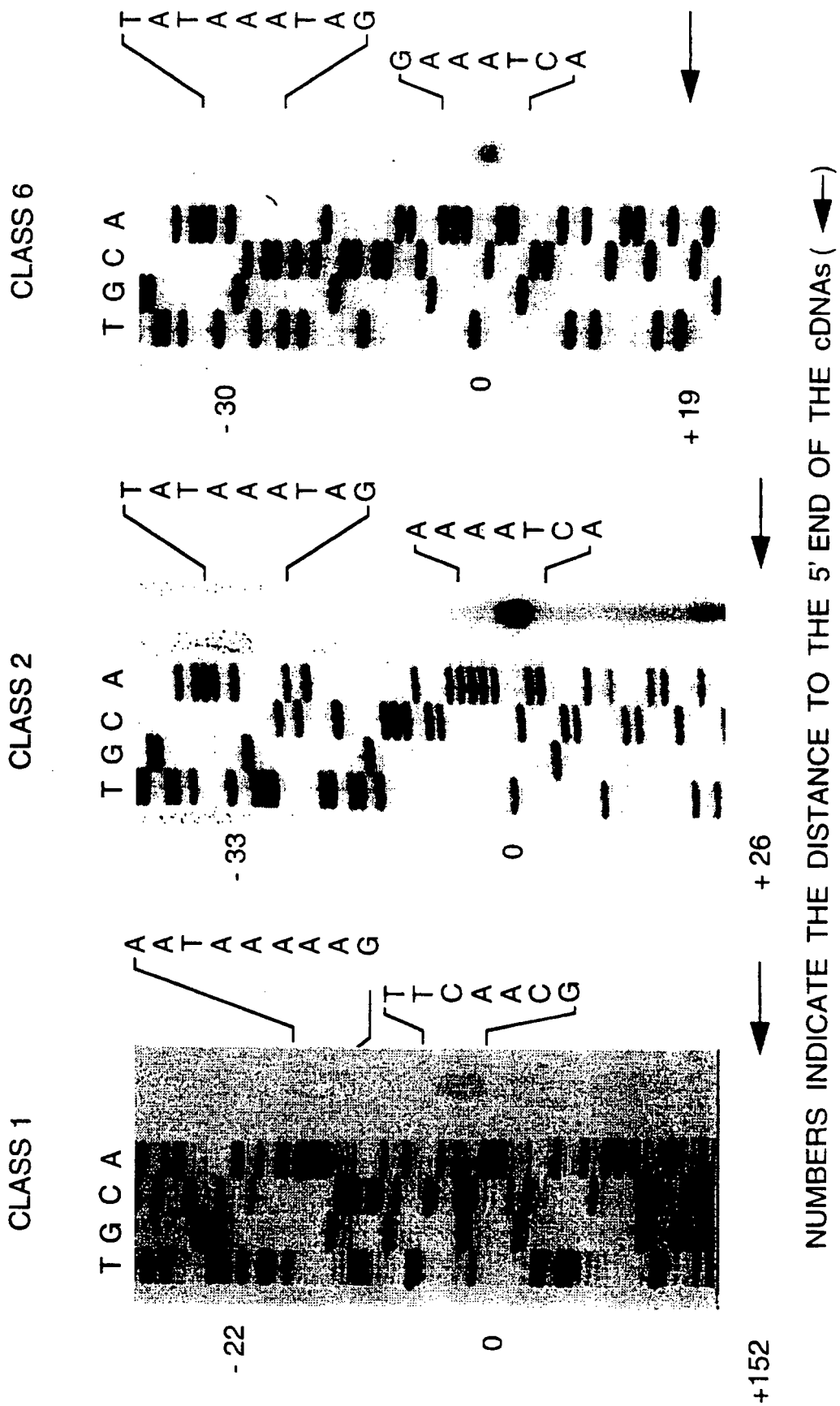
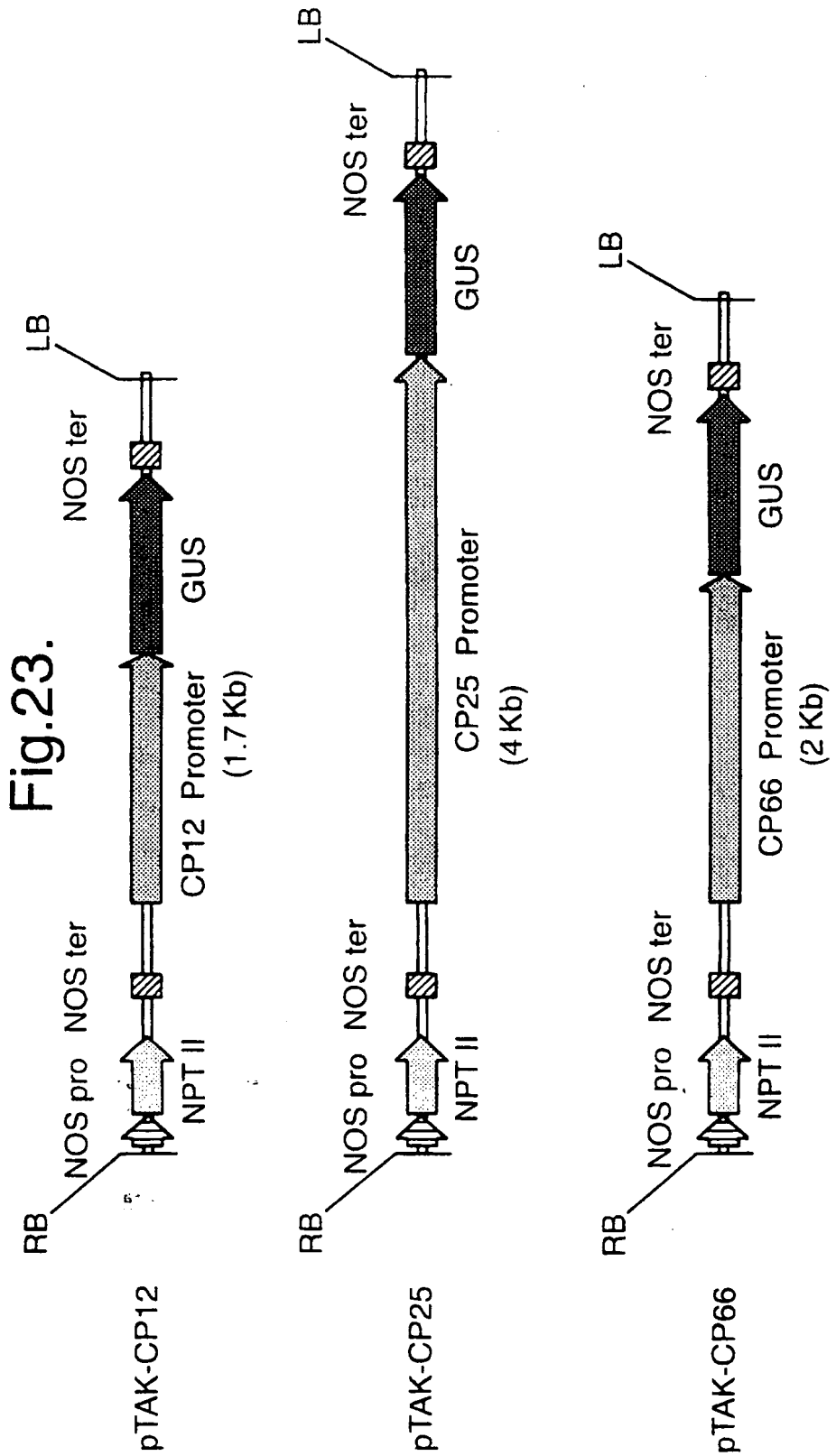

CLASS 1
CLASS 2
CLASS 6
T G C A
A A T A A A A G
T T C A A C G
A A A A T C A
T A T A A A T A G
G A A A T C A
- 22
0
+152
- 33
0
+ 26
- 30
0
+ 19
NUMBERS INDICATE THE DISTANCE TO THE 5' END OF THE cDNAs (←)

Fig.23.


LB
NOS ter
GUS
CP12 Promoter
(1.7 Kb)
NOS pro
NOS ter
NPT II
RB
pTAK-CP12
LB
NOS ter
GUS
CP25 Promoter
(4 Kb)
NOS pro
NOS ter
NPT II
RB
pTAK-CP25
LB
NOS ter
GUS
CP66 Promoter
(2 Kb)
NOS pro
NOS ter
NPT II
RB
pTAK-CP66

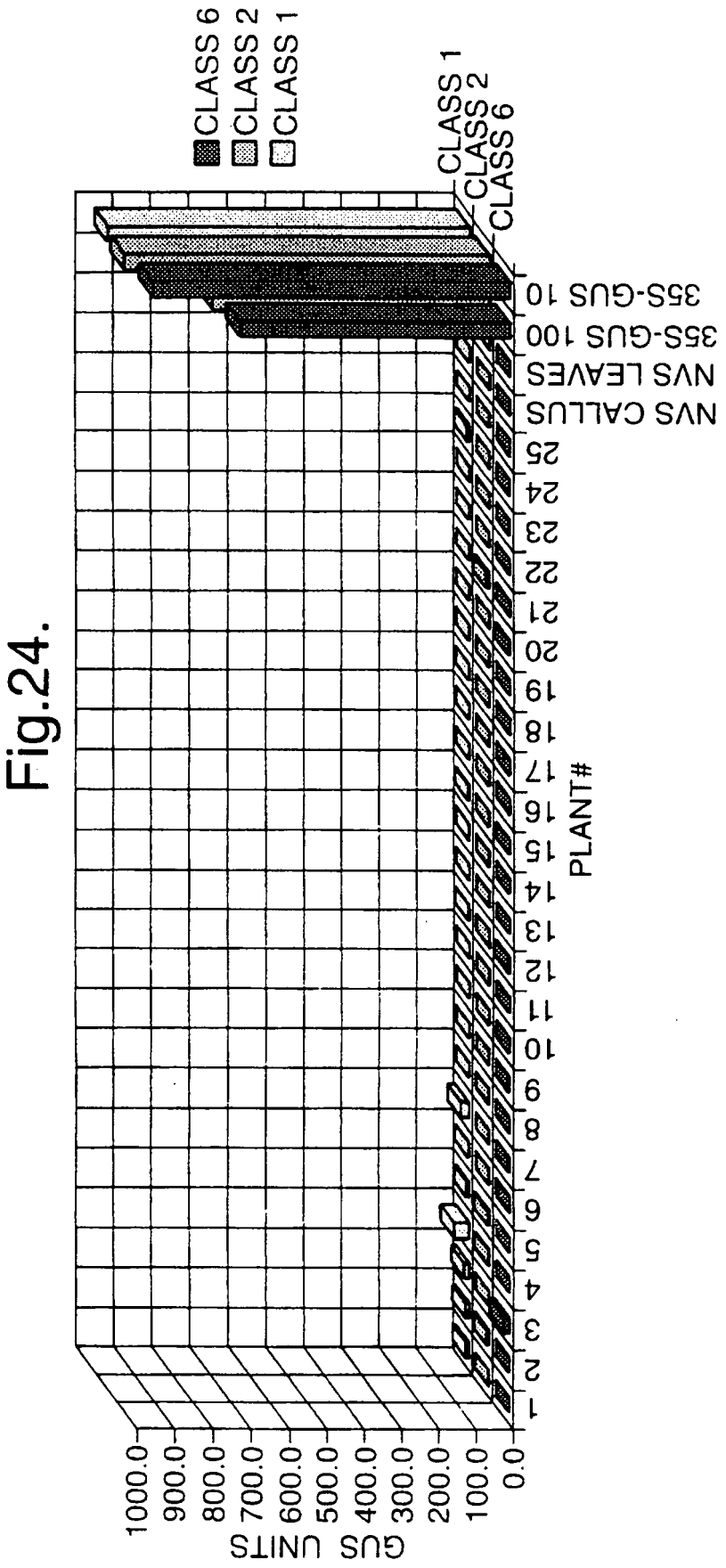

CLASS 6
CLASS 2
CLASS 1
CLASS 1
CLASS 2
CLASS 6
35S-GUS 10
35S-GUS 100
NVS LEAVES
NVS CALLUS
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
PLANT#
1000.0
900.0
800.0
700.0
600.0
500.0
400.0
300.0
200.0
100.0
0.0
GUS UNITS

Fig.24.

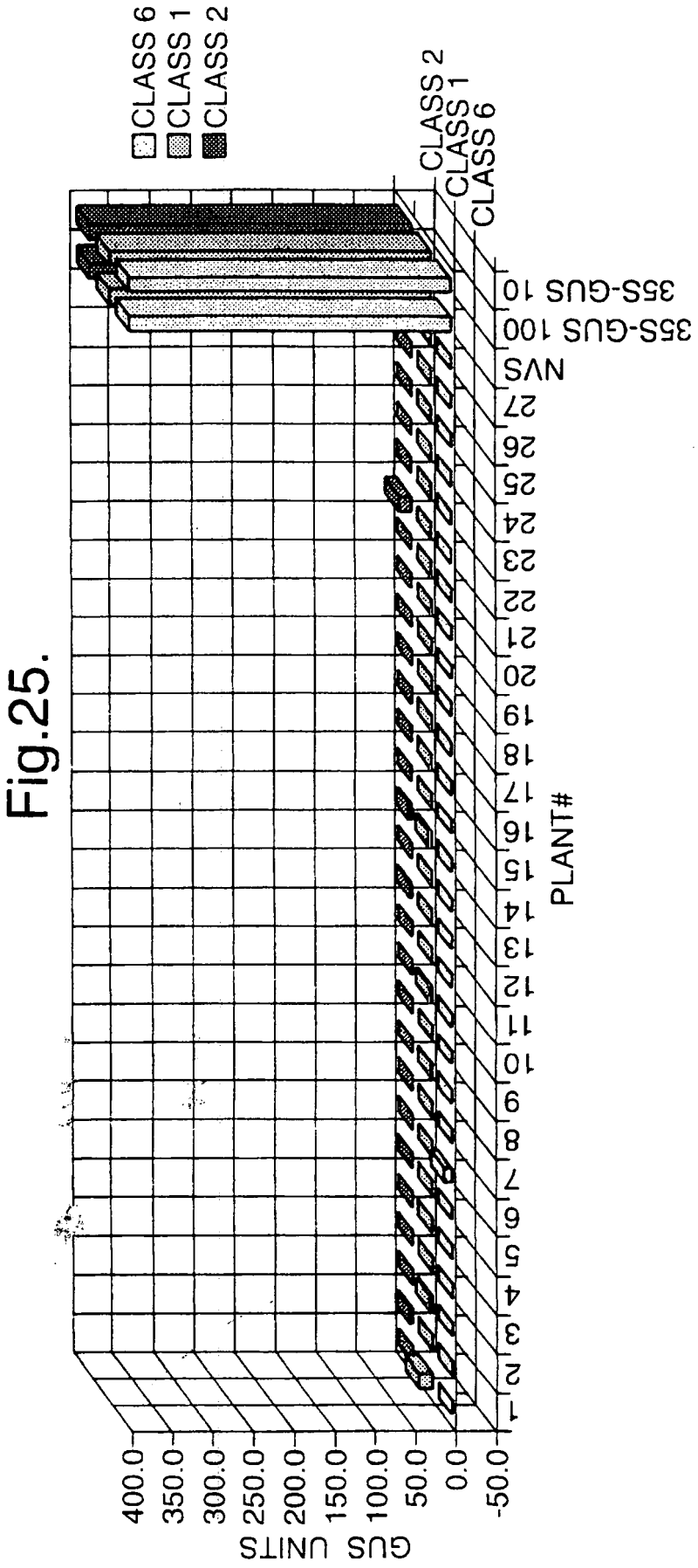

CLASS 6
CLASS 1
CLASS 2
CLASS 2
CLASS 1
CLASS 6
35S-GUS 10
35S-GUS 100
NVS
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
PLANT#
400.0
350.0
300.0
250.0
200.0
150.0
100.0
50.0
0.0
-50.0
GUS UNITS

Fig.25.

Fig.26.

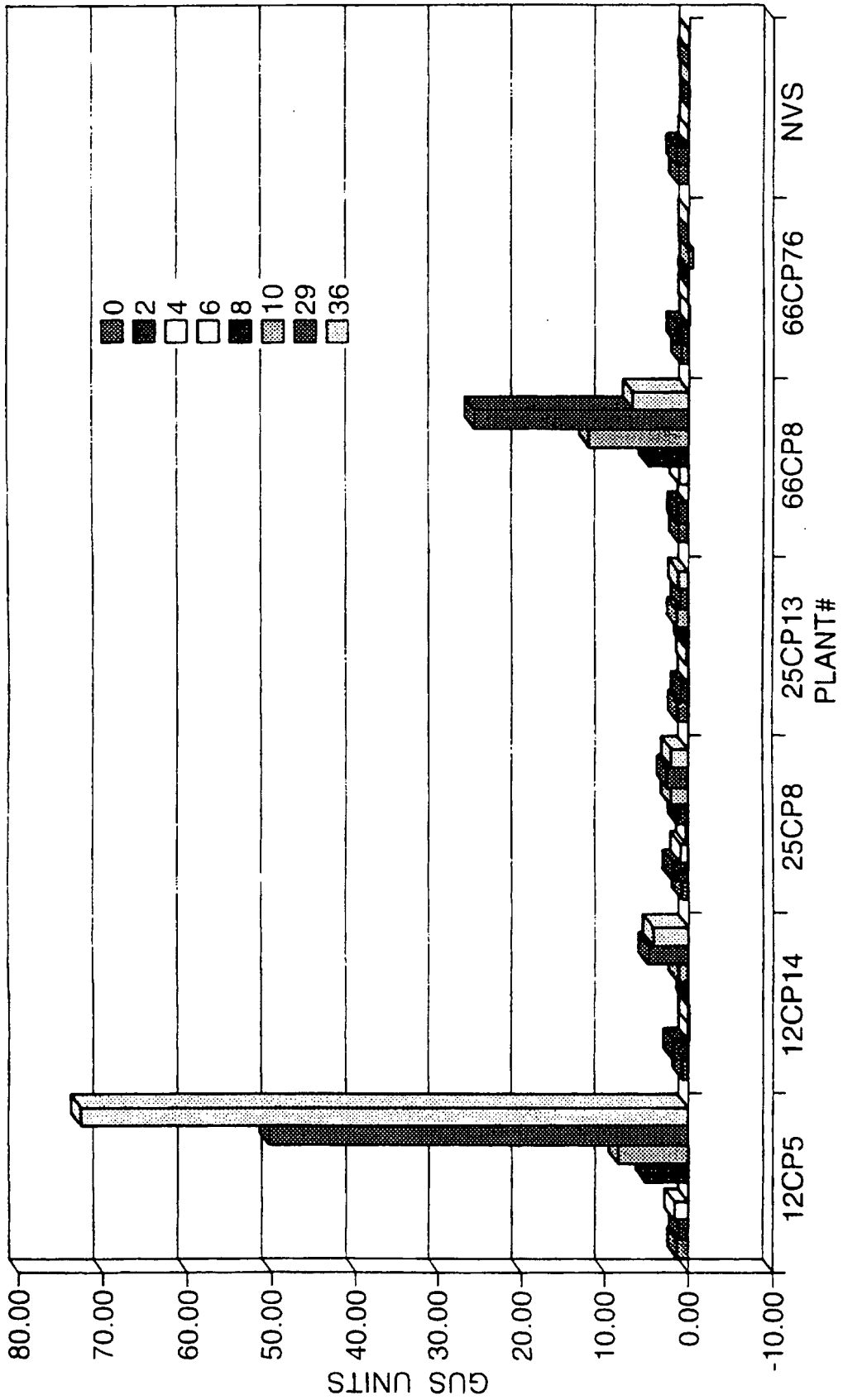

0
2
4
6
8
10
29
36
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
-10.00
GUS UNITS
12CP5
12CP14
25CP8
25CP13
66CP8
66CP76
NVS
PLANT#

Fig.27.

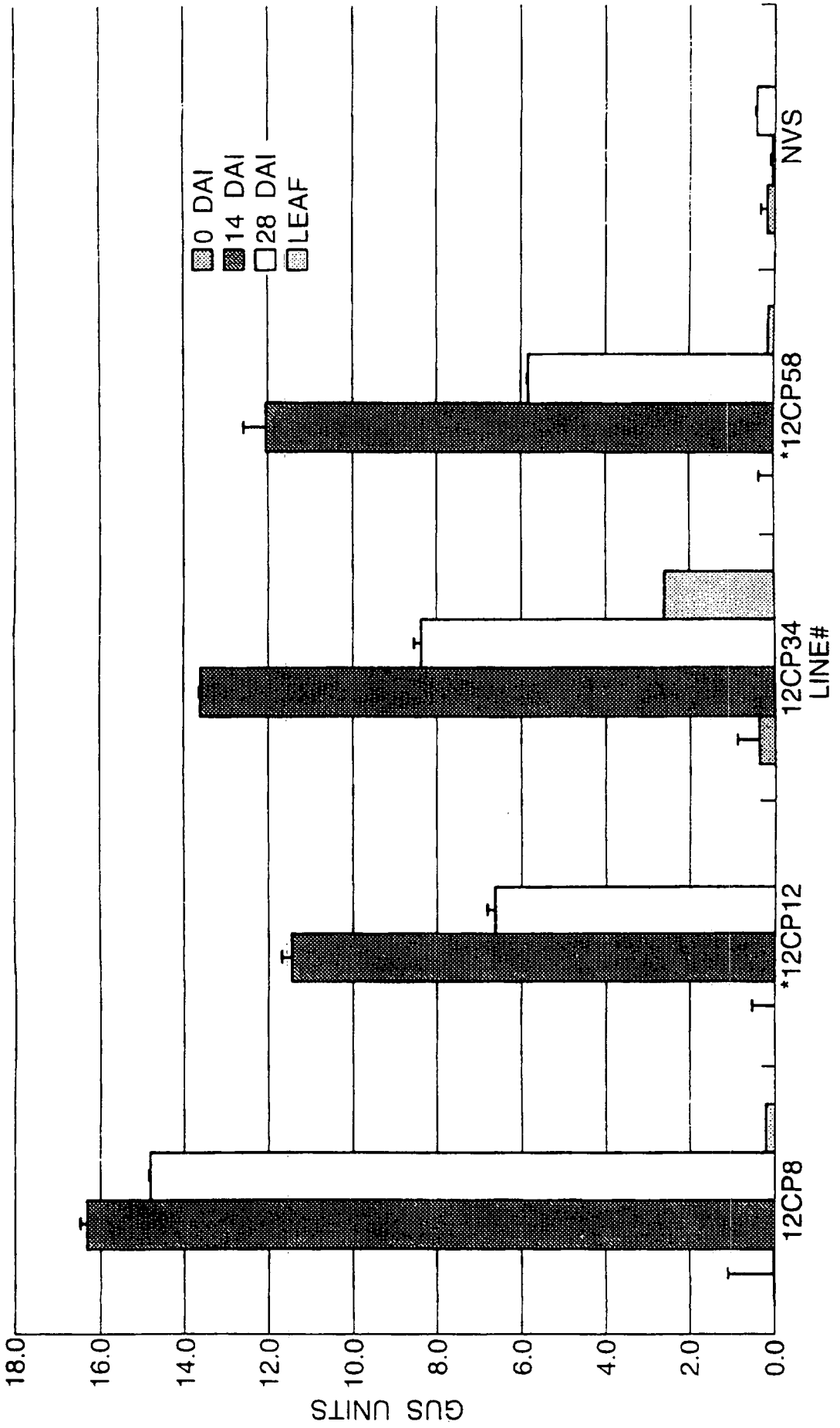
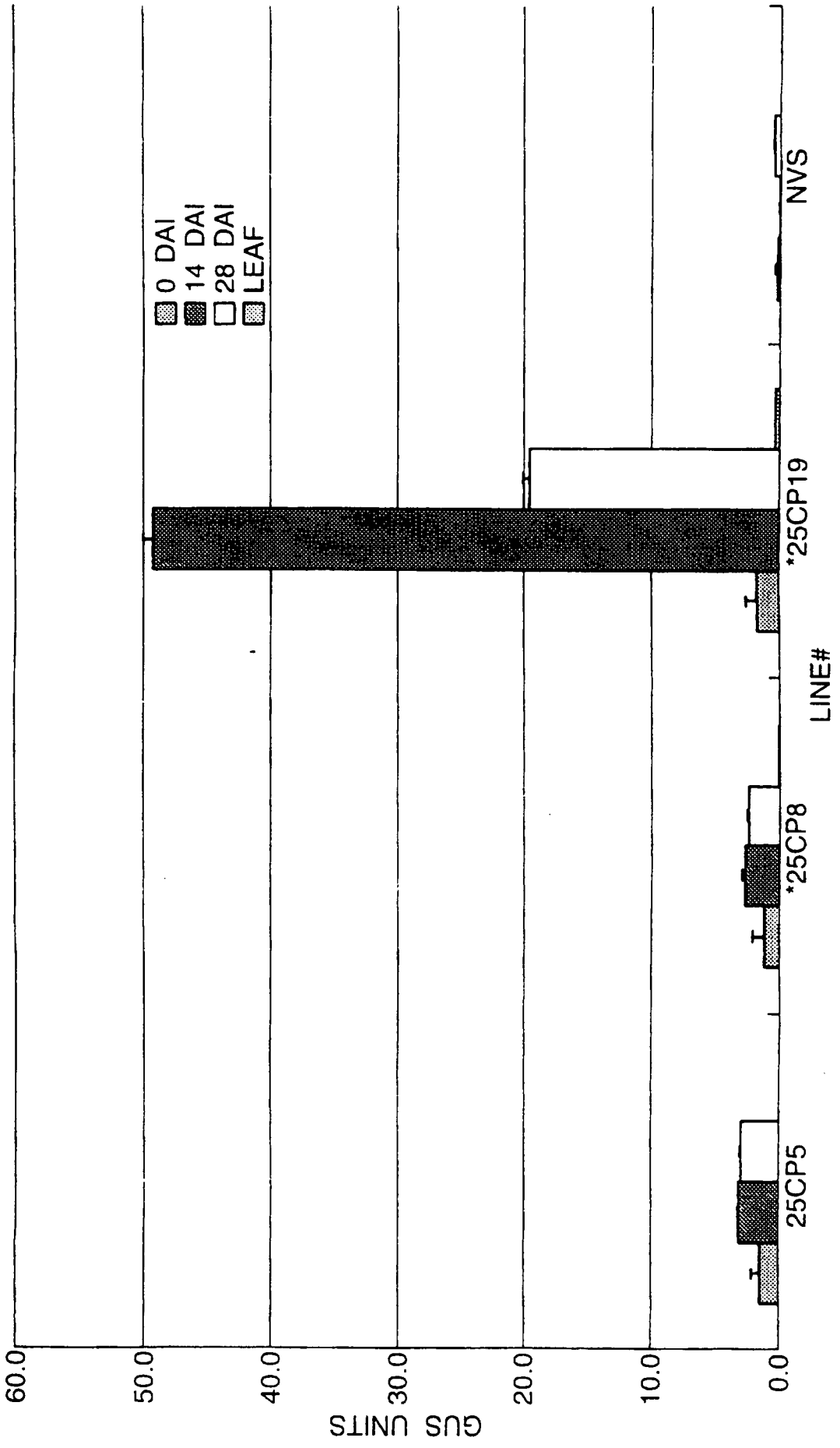

0 DAI
14 DAI
28 DAI
LEAF
GUS UNITS
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
12CP8
*12CP12
12CP34
*12CP58
NVS
LINE#

Fig.28.


0 DAI
14 DAI
28 DAI
LEAF
NVS
*25CP19
*25CP8
25CP5
LINE#
GUS UNITS
60.0
50.0
40.0
30.0
20.0
10.0
0.0

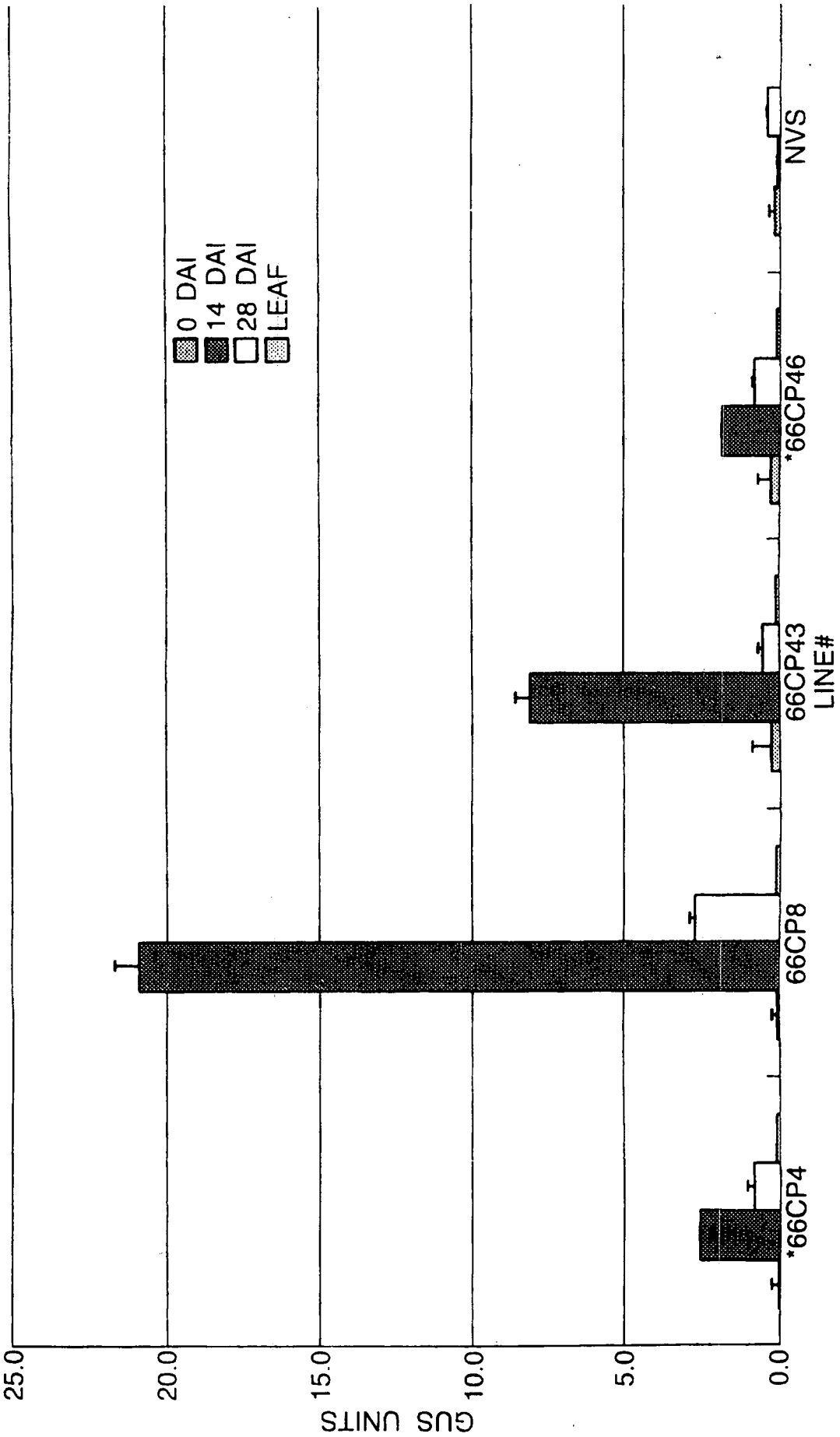

0 DAI
14 DAI
28 DAI
LEAF
25.0
20.0
15.0
10.0
5.0
0.0
GUS UNITS
*66CP4
66CP8
66CP43
*66CP46
NVS
LINE#

Fig.29.